ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 7.0.91— 2015 (ИСО 25964-1:2011)

Система стандартов по информации, библиотечному и издательскому делу

ТЕЗАУРУСЫ ДЛЯ ИНФОРМАЦИОННОГО ПОИСКА

(ISO 25964-1:2011, MOD)

Издание официальное

Москва
Стандартинформ
2016

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН Федеральным государственным бюджетным учреждением науки Всероссийским институтом научной и технической информации Российской академии наук (ВИНИТИ РАН) на основе собственного аутентичного перевода на русский язык англоязычной версии международного стандарта, указанного в пункте 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 191 «Научно-техническая информация, библиотечное и издательское дело»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 15 декабря 2015 г. № 2163-ст

4 Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО 25964-1:2011 «Информация и документация. Тезаурусы и их совместимость с другими словарями. Часть 1. Тезаурусы для информационного поиска» (ISO 25964-1:2011 «Information and documentation — Thesauri and interoperability with other vocabularies — Part 1. Thesauri for information retrieval», MOD). При этом дополнительные показатели и требования, включенные в текст стандарта для учета потребностей экономики Российской Федерации выделены подчеркиванием. В настоящий стандарт не включены графы таблицы 2, содержащие сокращения на немецком, датском, финском, норвежском, шведском и испанском языках, поскольку они не применяются в практике национальной стандартизации и информатики. Не включены также информационные приложения и предметный указатель. Поскольку вопросы построения многоязычных тезаурусов регулирует ГОСТ 7.24—2007, раздел 9 и подразделы 12.3 и 12.4 исключены из настоящего стандарта.

Наименование настоящего стандарта изменено относительно наименования указанного международного стандарта для приведения в соответствие с ГОСТ Р 1.5—2004 (пункт 3.5)

5 ВВЕДЕН ВПЕРВЫЕ

*Правила применения настоящего стандарта установлены в ГОСТ Р 1.0—2012 (раздел 8). Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© Стандартинформ, 2016

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 Термины и определения 1
3 Символы, сокращения и условные обозначения 9
4 Тезаурусы. Обзор и назначение 12
5 Понятия и объем их значения в тезаурусе 14
6 Термины тезауруса 17
7 Сложные понятия 30
8 Отношения эквивалентности в одноязычном контексте 37
9 Межъязыковая эквивалентность 42
10 Отношения понятий 42
11 Фасетный анализ 50
12 Представление и оформление тезауруса 51
13 Управление разработкой и ведением тезауруса 61
14 Указания по программному обеспечению для ведения тезауруса 70
15 Модель данных 74
16 Интеграция тезаурусов с приложениями 85
17 Обменные форматы 88
18 Протоколы 89
19 Библиография 91

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Система стандартов по информации, библиотечному и издательскому делу

ТЕЗАУРУСЫ ДЛЯ ИНФОРМАЦИОННОГО ПОИСКА

System of standards for information, librarianship and publishing.
Thesauri for information retrieval

Дата введения — 2016—07—01

1 Область применения

В настоящем стандарте установлены рекомендации, касающиеся развития и ведения информационно-поисковых тезаурусов. Эти рекомендации могут быть применены и к другим словарям, используемым в процессе информационного поиска во всех типах информационных ресурсов. При этом возможность такого применения не зависит от типа средств, используемых при передаче информации (текст, звук, неподвижный или движущийся образ, физический или мультимедийный объект), включая базы знаний, порталы, базы библиографических данных, тексты, музейные или мультимедийные коллекции в целом и входящие в их состав самостоятельные единицы.

В этом стандарте даются рекомендации, касающиеся развития и ведения информационно-поисковых тезаурусов. Эти рекомендации могут быть применены и к другим словарям, используемым в процессе информационного поиска во всех типах информационных ресурсов.

Сферой применения настоящего стандарта являются как одноязычные, так и многоязычные тезаурусы.

В сферу действия настоящего стандарта не входит подготовка указателей, помещаемых в конце изданий, однако некоторые из предлагаемых настоящим стандартом рекомендаций могут быть использованы и для этой цели.

Хотя рекомендации настоящего стандарта не предназначены для создания баз данных или программного обеспечения, используемого непосредственно в процессе поиска или индексирования, однако содержащиеся в данном стандарте рекомендации по вопросам тезаурусного менеджмента предполагают возможность разработки таких приложений.

2 Термины и определения

В настоящем стандарте применены следующие термины с соответствующими определениями:

2.1 **классификационный ряд** (array): Группа соподчиненных понятий (2.52).

Пример — Соподчиненные понятия «outerwear» (верхняя одежда) и «underwear» (нижняя одежда) вместе с понятием «clothing» (одежда) формируют классификационный ряд:

clothing	***одежда***
outerwear	***верхняя одежда***
overcoats	***пальто***
underwear	***нижняя одежда***

2.2 **ассоциативные отношения** (associative relationship): Отношения внутри пары понятий (2.11), которые не связаны иерархически, однако имеют сильную смысловую связь.

2.3 **вышестоящий термин** (broader term): Дескриптор (2.45), обозначающий более широкое понятие (2.11), чем обозначаемое данным дескриптором.

Примечание — Тематическая область более узкого понятия целиком находится внутри тематической области более широкого понятия. Отношения между этими двумя понятиями обычно обозначается меткой ВТ. Более подробные объяснения см. в 10.2.1.

Издание официальное

2.4 **основание деления** (characteristic of division): Признак, по которому понятие (2.11) может разделяться на ряд (2.1) более узких понятий (2.21), каждое из которых имеет значение этого признака, отличное от других.

Примечание — Ср. фасетный анализ (2.21), узловая метка (2.38).

Пример — Возрастная группа является основанием деления понятия «люди»:
люди
(по возрасту)
дети
молодежь
взрослые.

2.5 **классификация** (classification), **классифицирование** (classifying): Деятельность, подразумевающая объединение сходных и родственных объектов вместе; разъединение несходных и не имеющих родства объектов; и представление результирующих групп в логической и удобной последовательности.

2.6 **классификационная схема** (classification scheme): Таблица (2.49) понятий (2.11) и предкоординированных комбинаций понятий (2.11), организованная классификацией (2.5).

Примечание — В состав классификационной схемы часто входят также указатели понятий.

2.7 **термин-неологизм** (coined term): Новый термин (2.61), созданный для выражения понятия (2.11), для которого в соответствующем языке не существует подходящего термина (2.61).

Примечание — Более подробные объяснения см. в 6.6.5 и 8.3.3.3.

2.8 **сложная эквивалентность** (compound equivalence): Отношение или соответствие, в рамках которого один термин (2.61) или понятие (2.11) одного контекста представлено двумя или более терминами (2.61) или понятиями (2.11) в другом контексте.

2.9 **составной термин** (compound term): Термин (2.61), который в соответствие с морфологическими законами может быть разложен на самостоятельные отдельные компоненты.

Примеры —
1 В английском языке:
«Copper mine» можно расщепить на «coppe» и «mines»;
«lawnmowers» можно расщепить на «lawns» и «mower»
2 Во французском языке:
«mine de cuivre» можно расщепить на «mine» и «cuivre»;
«biodiversité» можно расщепить на «biologie» и «diversité».
3 В русском языке:
«медные шахты» можно расщепить на «медь» и «шахты»;
«биоразнообразие» можно расщепить на «биология» и «разнообразие».

Примечание — Составной термин может состоять как из нескольких слов, так и из одного слова.

2.10 **компьютерное приложение** (computer application): Компьютерная программа или набор компьютерных программ, с помощью которых осуществляется обработка данных высокого уровня в соответствии с определенными потребностями пользователя.

Примечание — В настоящем стандарте компьютерное приложение часто называют просто «приложение».

2.11 **понятие** (concept): Единица мысли.

Примечание — Понятия часто имеют несколько разных способов выражения. Они существуют в сознании в виде абстрактных единиц, которые независимы от терминов, используемых для их выражения. Они варьируют в широком диапазоне от очень простых понятий, например, «ребенок», до очень сложных, например, «законодательство о защите детей».

2.12 **контрольный словарь** (controlled vocabulary): Рекомендуемый список терминов (2.61), рубрик или кодов, каждый из которых представляет понятие (2.11).

Примечание — Контрольные словари проектируются для приложений, в которых полезно идентифицировать каждое понятие с помощью одной и той же (постоянной) рубрики, когда проводят классифицирование, индексирование и/или поиск документов.

2.13 **межъязыковая эквивалентность** (cross-language equivalence): Отношение эквивалентности (2.18) между терминами (2.61), представляющими одно и то же понятие (2.11) в различных языках.

2.14 **модель данных** (data model): Абстрактная модель, описывающая то, как данные представляются и используются.

Примечание — Модель данных в настоящем стандарте обеспечивает общее определение структуры и семантики тезауруса. Она может быть использована в качестве основы для определения либо модели базы данных, либо обменного формата тезаурусов.

2.15 **документ (document)**: Любой ресурс, который может быть классифицирован или индексирован для того, чтобы стал возможным поиск содержащихся в нем данных или информации.

Примечание — Это определение распространяется не только на материалы, написанные и отпечатанные на бумажном носителе или представленные в виде микрофильма (обычные книги, журналы, диаграммы, карты), но и на непечатные способы передачи информации. Например, такие как машиночитаемые носители и оцифрованные записи, ресурсы Интернета и интранета, фильмы, звукозаписи, люди и организации как носители научных знаний, здания (buildings), местности, монументы, трехмерные объекты действительности, а также собрания и составные части таких единиц.

2.16 **входной термин**, вводящий термин (entry term, lead-in term): Термин (2.61), представленный в контрольном словаре (2.12), но используемый не непосредственно в качестве метаданных (2.33), а для того, чтобы привести пользователя к другому термину (2.61), имеющему статус либо категориальной метки, либо предметного заголовка, либо дескриптора (2.45).

Примечание — Входной термин в составе тезауруса обычно трактуется как непредпочтительный термин, или аскриптор.

2.17 **эквивалентное отображение** (equivalence mapping): Соответствие, фиксирующее некоторое понятие (2.11) в целевом словаре (target vocabulary), которое признается идентичным по объему понятию (2.11) исходного словаря (source vocabulary).

Примечание — См. отношение эквивалентности (2.18).

2.18 **отношение эквивалентности** (equivalence relationship): Отношение между двумя терминами (2.61) в тезаурусе (2.62), показывающее, что оба эти термина обозначают одно и то же понятие (2.11).

Примечание — В обычном словоупотреблении это термины, являющиеся квазисинонимами и они могут представлять собой слегка различающиеся понятия. Однако включение в тезаурус устанавливаемого между ними отношения эквивалентности определяет, что оба эти термина рассматриваются как представители одного итого же понятия. Когда в одноязычный или многоязычный тезаурус включены два или более термина одного и того же языка, то один из них выбирается в качестве дескриптора, а другой в качестве аскриптора; а когда два или более таких терминов являются представителями разных языков в многоязычном тезаурусе, каждый из них может выступать как дескриптор в своем собственном языке, и эти отношения принято называть межъязыковой эквивалентностью.

2.19 **обменный формат** (exchange format): Машиночитаемый формат для представления информации, предназначенный для облегчения обмена информацией между различными приложениями.

Примечание — Обменный формат для тезауруса часто использует язык разметки, например, на основе стандарта XML (Extensible Markup Language) [63, 64, 65, 66] и основывается на модели данных тезауруса. Если модель данных представляет собой общее описание структуры и семантики тезауруса, то обменный формат выражает это на формальном языке для задачи обмена тезаурусами.

2.20 **фасет** (facet): Группа однородных понятий (2.11) одной и той же природной категории.

Примеры —

1 Животные, мыши, нарциссы и бактерии могут рассматриваться как члены фасета живых организмов.

2 Копание, писание и кипячение могут рассматриваться как члены фасета действий.

3 Париж, Великобритания и Альпы могут рассматриваться как члены фасета территорий.

Примечания

1 Примерами таких категорий высокого уровня, которые могут быть использованы для группировки понятий в фасеты, являются следующие категории: предмет, материал, действующий агент, действие, место и время.

2 Ср. узловая метка (2.38).

2.21 **фасетный анализ** (facet analysis): Выявление в предметной области входящих в ее состав понятий (2.11), сгруппированных в фасеты (2.20), и подразделение понятий (2.11) на более узкие понятия (2.11) на основе специальных оснований деления (2.4).

2.22 **фасетный индикатор** (facet indicator): Элемент классификационного индекса, который указывает начало нового фасета (2.20) внутри синтезированного сложного классификационного индекса (2.40).

Примечание — Примерами фасетного индикатора могут служить 0 в десятичной классификации Дьюи, скобки и кавычки в УДК. В прошлом термин «фасетный индикатор» использовался как синоним для термина «узловая метка», но во избежание путаницы в настоящем стандарте такое использование запрещено.

2.23 **иерархические отношения** (hierarchical relationship): Отношение между двумя понятиями (2.11), при котором объем одного из них полностью находится внутри объема другого.

Примечание — Существует несколько разных типов иерархических отношений. Более подробно об этом см. в 10.2. См. также вышестоящий (родовой) дескриптор (2.3), нижестоящий (видовой) дескриптор (2.37).

2.24 **омограф** (homograph): Одно, два или более слов, которые пишутся одинаково, но имеют разное значение.

Примеры —
1 В английском языке:
Слово «bank» может означать и «финансовый институт», и «берег реки».
2 Во французском языке:
Слово «avocat» может означать либо юриста, либо фрукт.
3 В русском языке:
Слово «лук» может означать либо растение, либо оружие.

Примечание — Омографы иногда называют омонимами, хотя последний термин имеет более широкое значение, поскольку включает амофоны, т.е. такие пары терминов как «weights» и «waits» в английском, «mer» и «mère» во французском или «код» и «кот» в русском языке, которые пишутся по-разному, а читаются одинаково.

2.25 **идентификатор** (identifier): Набор знаков, обычно алфавитно-цифровых, обозначающий понятие (2.11), термин (2.61) или какую-то другую сущность, используемый, особенно в компьютерных системах или сетях, для достижения однозначной идентификации внутри определенного контекста или ресурса.

Примечание — Иногда в качестве идентификатора используется классификационный индекс.

2.26 **индексный термин** (index term): Термин (2.61) приписанный документу (2.15) в процессе индексирования (2.27).

Примечание — Иногда индексные термины называют терминами индексирования, ключевыми словами или метками. Но два последних термина являются многозначными. В качестве индексных терминов часто используют дескрипторы тезаурусов.

2.27 **индексирование** (indexing): Интеллектуальный анализ предметного содержания документа (2.15) для идентификации представленных в нем понятий (2.11) и предоставление соответствующих индексных терминов (2.26) для обеспечения поиска информации.

Примечание — Для обозначения этого понятия используется термин «предметное индексирование (предметизация)», но поскольку в настоящем стандарте индексирование таких элементов как имена авторов, даты не рассматриваются, достаточно использовать термин «индексирование». Индексирование может осуществляться как пользователями-людьми, так и в автоматическом режиме.

2.28 **информационный поиск** (information retrieval): Все методы и процессы, используемые для того, чтобы выбрать из документной коллекции или сети информационных ресурсов документы (2.15), релевантные информационным потребностям.

Примечание — Это определение включает подбор и исключение документов из выборки, а также их просмотр и другие формы отыскания информации.

2.29 **совместимость** (interoperability): Способность двух или более систем или компьютеров обмениваться информацией и использовать информацию, полученную в результате такого обмена.

Примечание — Словари могут поддерживать совместимость путем включения связей с другими словарями, представления информации в стандартных форматах и путем использования систем, которые поддерживают общие компьютерные протоколы.

2.30 **заимствованный термин** (loan term): Термин (2.61), взятый из другого языка и принятый заимствующим языком.

Примеры —
1 «glasnost» — русский термин, который принят в английском языке;
2 «gourmet» — французский термин, который принят в английском языке;
3 «компьютер» — английский термин, который принят в русском языке.

2.31 **разметка** (markup): Примечания или какой-либо другой вид кодов, включенных в текст в соответствии с правилами языка разметки (2.32).

2.32 **язык разметки** (markup language): Набор правил кодирования, которые могут быть использованы для составления инструкций по интерпретации текста за счет использования примечаний, включенных непосредственно в сам текст.

Примечание — Интерпретация касается таких вопросов, как содержание, структура и представление текста. Широко используемые примеры включают HTML (Hipertext Markup Language) [59], который в основном касается представления, и XML (Extensible Markup Language) [61, 62, 63, 64] и указывает структуру текста.

2.33 **метаданные** (metadata): Данные, которые идентифицируют атрибуты документа (2.15), для которых типичным является поддержка функций размещения, доступа, документирования, оценки и/или выбора.

Примечание — Дескрипторы и классификационные индексы, выбранные в процессе индексирования, применяют в качестве значений метаданных.

2.34 **моноиерархическая структура** (monohierarchical structure): Иерархическая организация понятий (2.11) в тезаурусе (2.62) или классификационной схеме (2.6), в которой каждое понятие (2.11) может иметь непосредственно над собой только одно вышестоящее понятие (2.11).

Примечание — Ср. полииерархическая структура (2.42).

Пример — В моноиерархической структуре понятие «пианино» не может одновременно причисляться и к клавишным инструментам, и к струнным инструментам; чтобы определить его место в структуре следует выбрать одну из этих возможностей.

2.35 **многоязычный тезаурус** (multilingual thesaurus): Тезаурус (2.62), в котором термины (2.61) и структура отношений представлены на двух или более естественных языках.

2.36 **многословный термин** (multi-word term): Термин (2.61), состоящий более чем из одного слова.

Примечание — Ср. расщепляемый термин (2.9).

Пример — Cost benefit analysis, анализ цен и прибыли.

2.37 **нижестоящий (видовой) дескриптор** (narrower term): Дескриптор (2.45), представляющий **понятие** (2.11), которое по значению у́же, чем рассматриваемое понятие.

Примечание — Объем видового дескриптора полностью располагается внутри объема родового дескриптора. Отношение видового дескриптора к родовому обозначается меткой NT, а статус родового меткой BT. Более подробно см. 10.2.1.

2.38 **узловая метка, метка узла** (node label): Обозначение, проставляемое в иерархическом или классификационном указателе для того, чтобы показать, как упорядочены термины.

Примечание — Узловая метка не является ни дескриптором, ни аскриптором. Она включает один из двух видов информации:

a) имя фасета, к которому принадлежат следующие за ней термины;

b) атрибут или основание деления, с помощью которого отсортирован или сгруппирован классификационный ряд близкородственных терминов.

См. примеры в разделе 11.

2.39 **аскриптор** (non-preferred term, non-descriptor): Термин (2.61), который не приписывают документу (2.15), а используют лишь в качестве входа в тезаурус (2.62) или как указатель для нахождения заменяющего дескриптора.

Примечание — Ср. входной термин (2.16)

Пример —
hounds *пес*
USE dogs *см.* собака[1)]

Примечание — В этом примере «***hound***» и «*пес*»— аскрипторы, а «**dogs**» и «**собака**» — те дескрипторы, которые следует употреблять вместо них.

2.40 **классификационный код** (индекс) (notation, class code, class number, classmark): Набор знаков, представляющий **понятие** (2.11) в структурированном словаре (2.56), особенно в классификационной схеме (2.6).

Примеры —

Классификационный код	*Словарь-источник*	*Понятие*
07.04.4	*Тезаурус Международной организации труда (ILO)*	*Политика и развитие рыболовства*
622.342 2	*Десятичная классификация Дьюи*	*Добыча золота*
373.3.016:51	*Универсальная десятичная классификация*	*Курс математики в начальной школе*
SBS XEJ B	*Библиографическая классификация Блисса*	*Закон об охране исчезающих видов*
H40-H42	*Международная статистическая классификация болезней и сходных состояний здоровья*	*Глаукома*

Примечание — Классификационный индекс иногда используется для того, чтобы отсортировать или разместить понятия в предопределенном порядке и, по желанию, показать в указателе, каким образом структурированы и сгруппированы компоненты сложных понятий. Классификационный индекс может осуществлять связь между алфавитной и систематической частью тезауруса. В контексте классификационной схемы «понятия» часто именуют «темами», особенно, если они подобно приведенным выше примерам отличаются сложностью.

2.41 **парадигматические отношения** (paradigmatic relationship), априорные отношения (a priori relationship): Отношения между понятиями (2.11), которые им внутренне присущи.

Примечание — Такие отношения приводятся в структурированном словаре вне связи с каким-либо индексированным документом. Более подробное обсуждение вопроса о парадигматических и синтагматических отношениях приведено в 4.3.

2.42 **полииерархическая структура** (polyhierarchical structure): Такая иерархическая организация понятий (2.11) в тезаурусе (2.62) или классификационной схеме (2.6), при которой каждое понятие (2.11) может иметь более одного вышестоящего (родового) понятия (2.11).

Пример — В полииерархической структуре понятие «органы (музыкальные инструменты)» может одновременно причисляться и к клавишным инструментам, и к духовым инструментам.

Примечания

1 Ср. моноиерархическая структура (2.35).

2 В полииерархической структуре единичное понятие может появляться более чем в одном месте иерархической структуры тезауруса. Его атрибуты и связи и особенно нижестоящие и вышестоящие термины остаются неизменными вне зависимости от того, где термин встретился.

2.43 **посткоординация** (post-coordination): Комбинирование дескрипторов (2.45) из контрольного словаря (2.12), осуществляемое во время поиска.

Примечание — Ср. предкоординация (2.44).

Пример — Посткоординированное поисковое предписание «микроволны AND излучения» может использоваться, чтобы получить документы о микроволновом излучении, когда они были проиндексированы с помощью отдельных терминов «микроволны» и «излучения», а не с помощью объединенного термина.

---

1) Различие шрифтов дескриптора и аскриптора в этом примере соответствует рекомендациям по представлению их в тезаурусе.

2.44 **предкоординация** (pre-coordination): Комбинирование понятий (2.11), классов или терминов (2.61) из контрольного словаря (2.12) во время создания этого словаря или во время использования его для индексирования (2.27) или классифицирования (2.5).

Примечание — Ср. посткоординация (2.43).

Примеры —

1 Класс «Общая теория», когда он находится в составе более широкого класса «музыка», соотносится только с предкоординированной темой «теория музыки», а не с теорией вообще.

2 Предкоординированная цепочка «картон — переработка» может появиться в словаре предметных рубрик или, если она не была туда включена, то она может быть синтезирована индексатором, если окажется необходимой для индексирования конкретного документа.

2.45 **дескриптор** (preferred term, descriptor): Термин (2.61), используемый в тезаурусе для представления понятия (2.11) при индексировании (2.27)[1)].

Примечания

1 Ср. аскриптор (2.39).

2 Дескриптор — это, как правило, существительное или субстантивное словосочетание.

2.46 **протокол** (protocol): Соглашение, которое определяет синтаксис, семантику и синхронизацию процесса коммуникации между двумя компьютерами для обеспечения информационного поиска.

2.47 **квазисиноним,** неполный синоним (quasi-synonym, near-synonym): Один из двух или более терминов (2.61), значения которых в рамках обычного использования, как правило, рассматриваются как различные, но которые в данном контрольном словаре (2.12) могут рассматриваться в качестве меток для одного и того же понятия (2.11).

Примеры —

diseases, disorders	***болезни, недомогания***
earthquakes, earth tremors	***землетрясения, сейсмические события***

2.48 **ассоциативный термин** (related term): Дескриптор (2.45), обозначающий такое понятие (2.11), которое имеет ассоциативную связь (2.2) с рассматриваемым термином.

Примечание — Отношения между ассоциативными терминами обычно обозначаются меткой RT. Более подробные объяснения см. в 10.3.

2.49 **классификационная таблица** (schedule): Совокупность терминов (2.61), классификационных кодов (2.40), заголовков, перекрестных ссылок и лексических примечаний (2.50), которая служит для представления содержания и структуры структурированного словаря (2.56).

2.50 **лексическое примечание** (scope note): Запись, которая определяет или уточняет семантические границы понятия (2.11) в рамках его использования в **структурированном словаре** (2.56).

Примечание — Термин, используемый как метка для обозначения понятия, при обычном использовании имеет несколько значений. Лексическое примечание используют для закрепления за ним только одного из таких значений и, где это необходимо, оно отсылает к другим понятиям, которые включены или исключены из объема уточняемого понятия.

2.51 **поисковый термин** (search term): Термин (2.61), формирующий поисковый запрос или его часть.

Примечание — В контексте настоящего стандарта поисковые термины обычно выбираются из контролируемого словаря.

2.52 **соподчиненное понятие** (sibling concept): Одно из двух или более понятий (2.11) с одним и тем же ближайшим вышестоящим понятием (2.11), каждое из которых представлено дескриптором (2.45).

[1)] Более точное определение дано в [5] «Дескриптор: Лексическая единица, выраженная информативным словом (вербально) или кодом и являющийся именем класса синонимичных или близких по смыслу ключевых слов.» В контексте данного стандарта дескриптор отличается от индексного термина тем, что дескриптор рассматривается как компонент тезауруса.

Пример — Outerwear (верхняя одежда) и underwear (нижняя одежда) являются дескрипторами, которые обозначают соподчиненные понятия, принадлежащие к одному и тому же классификационному ряду:

clothing	***одежда***
outerwear	***верхняя одежда***
overcoats	***пальто***
underwear	***нижняя одежда.***

2.53 **соподчиненный термин** (sibling term): Один из двух или более дескрипторов (2.45), имеющих один и тот же ближайший вышестоящий (родовой) термин (2.3).

Пример — Дескрипторы chairs (сидения) и tables (столы), являются соподчиненными терминами одного и того же классификационного ряда, в то время как дескрипторы «furniture (мебель), armchairs (кресла) и dining tables (обеденные столы») соподчиненными терминами не являются:

furniture	***мебель***
chairs	***сиденья***
armchairs	***кресла***
tables	***столы***
dining tables	***обеденные столы***

2.54 **исходный язык *(source language)***: Язык, служащий в качестве отправной точки в процессе перевода или поиска эквивалентов для терминов (2.61).

2.55 **специфичность словаря** (specificity): Способность структурированного словаря (2.56) выразить предмет поиска углубленно и подробно.

Примечание — Более подробное объяснение см. обсуждение специфичности в 8.4 и других местах.

2.56 **структурированный словарь** (structured vocabulary): Организованный набор **терминов** (2.61), заголовков и кодов, представляющих понятия (2.11) и их взаимосвязи, которые могут быть использованы для поддержки информационного поиска (2.28).

Примечание — Структурированный словарь также может быть использован для других целей. В рамках поиска информации словарь нуждается в сопутствующих правилах, описывающих как следует применять термины. Вопросы различных типов структурированных словарей, в том числе классификационных схем, словарей предметных рубрик и др., будут переадресованы к настоящему стандарту.

2.57 **словарь предметных рубрик,** язык предметных рубрик (subject heading scheme, subject heading language, subject heading list, SHL): Структурированный словарь (2.56), состоящий из терминов (2.61), доступных для предметного индексирования (2.27), плюс правила для объединения их в предкоординированные цепочки терминов (2.61) для индексирования.

2.58 **синоним** (synonym): Один из двух или более терминов (2.61), обозначающих одно и то же понятие (2.11).

Примеры —

В английском языке:

guarantees, warranties
heart attack, myocardial infarction
HIV, human immunodeficiency virus

Во французском языке:

schiste, phyllade
VIH, virus de l'immunodéficience humaine
crise cardiaque, infarctus du myocarde

В русском языке:

чахотка, туберкулез
ВИЧ, вирус иммунодефицита человека
кровоизлияние в мозг, инсульт.

Примечание — Сокращение и полная формы термина могут рассматриваться как синонимы.

2.59 **синтагматические отношения,** апостериорные отношения (syntagmatic relationship, a posteriori relationship): Такие отношения между понятиями (2.11), которые существуют только потому, что эти понятия встретились совместно в индексируемом документе (2.15).

Примечание — Вне рамок индексируемого документа такие отношения, как правило, не действуют, поэтому они не входят в структуру тезауруса. Для ознакомления с более полным обсуждением синтагматических и парадигматических отношений см. 4.3.

2.60 **целевой язык** (target language): Язык, в котором находится перевод или эквивалент термина (2.61) исходного языка (2.54).

2.61 **термин** (term): Слово или словосочетание, используемое для обозначения **понятия** (2.11).

Примеры —

Schools	***школы***
school uniform	***школьная форма***
costs of schooling	***плата за школу***
teaching	***преподавание.***

Примечание — Термины тезауруса могут быть либо дескрипторами, либо аскрипторами.

2.62 **тезаурус** (thesaurus): Контрольный (2.12) структурированный словарь (2.56), в котором понятия (2.11) представлены терминами (2.61), организованными таким образом, что отношения между понятиями (2.11) представлены эксплицитно, и дескрипторы (2.45) снабжены указателями перехода от синонимов (2.58) и квазисинонимов (2.47).

Примечание — Задачей, решаемой тезаурусом, является обеспечение того, чтобы как индексатор, так и пользователь выбирали для представления данного содержания (предмета) один и тот же дескриптор или комбинацию дескрипторов. По этой причине тезаурус оптимизирован так, чтобы стать средством навигации и терминологического покрытия предметной области для человека.

2.63 **наивысший термин** (top term, TT): Дескриптор (2.45), представляющий понятие (2.11), для которого в тезаурусе (2.62) не существует более широкого понятия.

2.64 **управление словарем** (vocabulary control): Словарная работа, проводимая для того, чтобы избежать многозначности и упорядочить форму представления терминов (2.61), а также лимитировать число понятий (2.11) и терминов (2.61), допустимых для использования в процессе индексирования (2.27).

Примечание — Цель управления достигается путем различения омографов так, чтобы каждый из них имел только одно значение, и путем выбора из ряда синонимов или квазисинонимов одного, рекомендуемого для использования при индексировании. Цель этих ограничений состоит в повышении вероятности того, что индексатор при индексировании и пользователь при поиске выберут для обозначения определенного понятия один и тот же термин.

3 Символы, сокращения и условные обозначения

3.1 В таблице 1 приведены сокращения, которые используются в англоязычных примерах тезаурусных записей в качестве меток к префиксным терминам и записям. Каждая метка указывает на отношение или функцию термина или записи, следующую за ними.

3.2 В тезаурусах на голландском языке без перевода используются те же метки, что и в тезаурусах на английском языке.

3.3 Сокращения, приведенные в таблице 2 в качестве меток, обрели статус общепризнанных соглашений и встречаются во многих опубликованных тезаурусах. Они имеют мнемоническое значение, но применяются только в отдельных языках.

В случае использования нейтральной системы аббревиатур рекомендуется использовать символы из левого столбца таблицы 1.

Таблица 1 — Символы и сокращения

Описание	Символ	Метка	Значение
Описательные элементы		SN	Лексическое примечание
		DEF	Определение
		HN	Историческая справка

Окончание таблицы 1

Описание	Символ	Метка	Значение
Коды		SC	Предметная область. Классификационный код или индекс группы понятий, относящихся к данной теме
		CC	Классификационный код или индекс понятия
Отношения	→	USE	«Используй». Термин, стоящий после этой метки, является дескриптором, которым надо заменить аскриптор, предшествующий этой метке
	=	UF	«Используй вместо» или «Используется вместо». Термин, стоящий после этой метки, является аскриптором, для которого предшествующий этой метке термин является дескриптором, употребляемым вместо этого аскриптора
		USE...+	Два или более дескриптора, стоящие после этой метки, следует использовать вместе для того, чтобы представить понятие, предшествующее этой метке
		UF...+	Аскриптор, стоящий после этой метки, должен быть представлен комбинацией дескрипторов, один из которых предшествует этой метке
		TT	Наивысший термин. Предшествующий этой метке дескриптор является самым широким по объему понятием в иерархии, к которой принадлежит данное понятие
	<	BT	Вышестоящий термин. Стоящий после этой метки термин представляет понятие с более широким значением
		BTG	Выше — род. Более широкий термин, обозначающий класс (род), в который входит данное видовое понятие
		BTI	Выше — множество. Более широкий термин, обозначающий множество, в которое входит данное понятие
		BTP	Выше — целое. Более широкий термин, обозначающий целое, в которое входит данное понятие в качестве составной части
		NT	Нижестоящий термин. Стоящий после этой метки термин представляет понятие с более конкретным значением
		NTG	Ниже — вид. Более узкий термин, обозначающий подкласс (вид) данного родового понятия
		NTI	Ниже — элемент. Более узкий термин, обозначающий элемент множества, представленного данным понятием
		NTP	Ниже — часть. Более узкий термин, обозначающий составную часть объекта, представленного данным понятием
		RT	Ассоциативный термин. Следующий за этой меткой термин является ассоциативным термином, но не синонимом, не квазисинонимом, не вышестоящим и не нижестоящим термином

Таблица 2 — Метки на английском языке и их эквиваленты на других языках

Метка на английском языке	Метка на французском языке	Метка на китайском языке	Метка на русском языке
SN Scope note	NE Note explicative [b] Note d'emploi	J 解 jie	**лп** Лексическое примечание
USE Use	EM Employer	Y 用 yong	**см** Смотри

Окончание таблицы 2

Метка на английском языке	Метка на французском языке	Метка на китайском языке	Метка на русском языке
UF Use for *Used for* [a]	EP Employer pour *Employé pour* [a]	D 代 dai	**с** Синоним
BT Broader term	TG Terme générique	S 属 shu	**в** Выше
NT Narrower term	TS Terme specifique	F 分 fen	**н** Ниже
RT Related term	TS Terme associé	C 参 can	**а** Ассоциация
BTG Broader term (generic)	TG Terme générique (générique)		**вр** Выше — род
BTP Broader term (partitive)	TG Terme générique (partitif)		**вц** Выше — целое
BTI Broader term (instantial)	TG Terme générique (instance)		**вм** Выше — множество
NT Narrower term (generic)	TS Terme specifique (générique)		**нв** Ниже — вид
NT Narrower term (partitive)	TS Terme specifique (partitif)		**нч** Ниже — часть
NT Narrower term (instantial)	TS Terme specifique (instance)		**нэ** Ниже — элемент
Примечание — Набор языков в таблице 2 — открытый. Следующие издания этой части ИСО 25964 могут включать дополнительные метки. Для использования в национальном стандарте России из таблицы 2 исключены языки, не являющиеся официальными языками ИСО.			
[a] Строки курсивом показывают варианты расшифровки меток. [b] Французский стандарт AFNOR Z47-100 предлагает метку NA — Note d'application.			

3.4 В настоящем стандарте следующие соглашения также используются. Следует обратить внимание, что эти соглашения не являются обязательными для всех тезаурусов.

a) Распределение заглавных и прописных букв в дескрипторах совпадает с тем, как это имеет место в полном печатном тексте. Как правило, это означает, что нижний регистр используется во всех случаях, кроме случая использования верхнего регистра в составе сокращений и начальных букв в собственных именах.

Примеры [1)]

animals ***животные***
cars ***автомобили***
World Health Organization ***Всемирная организация здравоохранения***

b) Аскрипторы набирают курсивом.

Примеры —

animals ***UF fauna***	***fauna*** ***USE animals***
cars ***UF automobiles***	***automobiles*** ***USE cars***

[1)] В настоящих примерах прямой и курсивный шрифты терминов применяются в соответствии с рекомендациями настоящего стандарта по представлению дескрипторов и аскрипторов в тезаурусах.

World Health Organization *UF WHO*	*WHO* *USE World Health Organization*
животные *с фауна*	*фауна* *см animals*
автомобили *с автомашины*	*автомашины* *см автомобили*
Всемирная организация здравоохранения *с ВОЗ*	*ВОЗ* *см Всемирная организация здравоохранения*

4 Тезаурусы. Обзор и назначение

4.1 Общее назначение

Основным назначением тезауруса является обеспечение того, чтобы как при индексировании, так и при поиске были выбраны в качестве отражения одного и того же понятия одни и те же термины. Для достижения этой цели в тезаурусе, во-первых, должны быть заданы все понятия, которые могут быть полезны для обеспечения поиска в данной области. Понятия представляют с помощью терминов, и для каждого понятия одно из таких возможных представлений выбирают в качестве дескриптора (см. 4.2 и 6.6). Во-вторых, тезаурус должен представлять дескрипторы таким образом, чтобы пользователь мог легко идентифицировать тот дескриптор, который ему нужен. Это достигается путем задания отношений между терминами и/или понятиями (см. разделы с 8 по 10) и путем использования этих отношений при создании структурированных указателей терминов.

При включении в системы поиска или индексирования терминов и отношений тезауруса они могут быть использованы по-разному, в частности:

- как средство расширения поиска;
- для предложения альтернативных поисковых терминов;
- для поддержки кластеризации результатов или других средств совершенствования поиска;
- для выявления типичных орфографических ошибок;
- для поддержки автоматического индексирования.

Взаимосвязь терминов, понятий и отношений иллюстрируется на модели данных в разделе 15. Из практических соображений в этой части ИСО 25964 обычно идет речь о «терминах». Но никогда не следует забывать, что целью манипулирования терминами является поддержка поиска обозначаемых ими понятий.

4.2 Словарный контроль и его назначение

Процесс принятия решения о предоставлении термину права представлять определенное понятие не всегда является однозначно простым, поскольку понятие часто может быть выражено не одним способом. Более того, в обычном словоупотреблении некоторые термины могут иметь более одного значения. Поэтому словарный контроль является существенным моментом, и тезаурусы применяют для обеспечения однозначности путем использования следующих двух принципиально важных приемов.

а) Сфера применения понятий и терминов намеренно ограничивают предписанными значениями. В отличие от терминов таких словарей, в которых слова даются в сопровождении целого ряда различных определений, отражающих нормативное использование, каждый термин в тезаурусе, как правило, бывает привязан к тому или иному единственному значению, которое наиболее эффективно отражает потребности поисковой системы. Структура тезауруса, особенно представление в указателе иерархических отношений, часто указывает на нужное значение термина. Если это ограничение не достаточно очевидно, термин должен быть снабжен лексическим примечанием. Это примечание должно уточнить значение выбранного термина и может также указывать на другие значения, известные в естественном языке, которые были сознательно исключены для целей информационного поиска.

б) Если в одном и том же языке одно и то же понятие может быть выражено двумя или более синонимами или квазисинонимами, то в качестве дескриптора, как правило, выбирают один из этих терминов, который затем последовательно используют в качестве дескриптора в процессе индексирования

как главный или единственный термин, выражающий соответствующее понятие. При этом от любого синонима, который может понадобиться пользователю при обращении к тезаурусу, должна быть дана ссылка на заменяющий его дескриптор.

При осуществлении словарного контроля одним из последствий использования мер, описанных в перечислениях a) и b) является то, что полученные выражения могут не соответствовать. Тезаурус играет важную роль в посредничестве между терминами, используемыми в обычном словоупотреблении, и теми терминами, которые эффективно функционируют в процессе поиска информации. Для достижения эффективности поиска пользователи должны использовать с определенной степенью искусственности контрольного словаря (хотя в некоторых системах эта трудность может быть преодолена за счет автоматической замены термина пользователя дескриптором).

Для того чтобы тезаурус мог эффективно функционировать в многоязычном контексте, включаемые понятия должны быть представлены во всех используемых языках, обеспечивая доступ носителям этих языков. Если тезаурус является «симметричным», как описано в настоящем стандарте и смоделировано в разделе 15, то для каждого понятия должен существовать дескриптор в каждом из языков, и объем понятий должен быть одинаковым во всех языках. Это ограничение иногда вносит свой вклад в искусственность языка. Альтернативные подходы (например, использование метода сопоставления словарей) будут описаны в следующей версии стандарта [6].

4.3 Парадигматические и синтагматические отношения

Целью установления отношений и отображения их в указателе является указание пользователю (или специалисту, действующему от его имени) нужного направления для выбора наиболее подходящего термина, выражающего данное понятие. Это достигается путем демонстрации таких терминов, которые пользователь мог бы выбрать вместо или наряду с терминами, выбранными им изначально. Примеры приведены в разделе 12.

В любом языке между терминами индексирования, как правило, наблюдаются отношения следующих двух видов:

a) отношения, называемые синтагматическими, фиксируются в том случае, когда понятия встречаются вместе в контексте конкретного документа. Иными словами, если два или более термина индексирования приписаны одному и тому же документу, то тем самым между ними установлено синтагматическое отношение. Однако включение синтагматических отношений в тезаурус не рекомендуется.

Пример — Индексатор работы о «компьютерах в амстердамских банках» может выделить три понятия, выражаемые, соответственно, терминами «банки (финансовые учреждения)», «компьютеры» и «Амстердам». Он присвоит их этому документу. В посткоординированной системе связь между этими терминами в метаданных эксплицитно не указывается, но документ, тем не менее, может быть найден, если любой или все эти термины были использованы в качестве поисковых ключей. В предкоординированном указателе любая комбинация этих трех терминов может быть приведена вместе со ссылкой на то место, где соответствующий документ может быть найден. Понятия из данного примера в рамках обыденных отношений, как правило, никак друг с другом не связаны, поэтому их взаимосвязь должна рассматриваться как зависящая от документа.

b) парадигматическими называют отношения, которые значимы почти во всех контекстах, особенно если они внутренне присущи понятиям, которые представлены этими терминами. Наличие парадигматических отношений между тезаурусными понятиями продемонстрировать целесообразно, так как они часто ведут пользователей к понятиям, тесно связанным с теми терминами, которые пришли в голову первоначально.

Пример — Независимо от темы индексируемого документа понятия из предыдущего примера имеют такие сущностные связи: «банки (финансовые учреждения)» с более широким понятием «финансовые институты»; «компьютеры» устойчиво ассоциируют связь с понятием «обработка данных», а понятие «Амстердам» неразрывно связано с понятием «Нидерланды». Найдя в словаре любой из этих связанных между собой терминов, пользователь может быть заинтересован в информации, заиндексированной с помощью связанного термина. Такие парадигматические отношения не зависят от какого-либо конкретного документа. Они, как правило, распознаются и могут быть установлены благодаря ссылке на такие обычные издания как словари и энциклопедии.

Различие между этими двумя видами отношений может быть отображено так, как показано на рисунке 1.

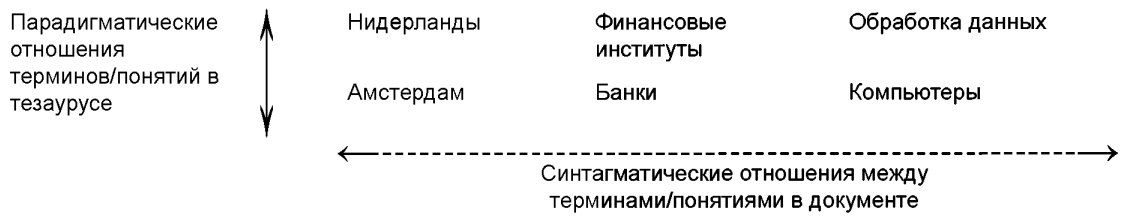


Рисунок 1 — Парадигматические и синтагматические отношения

4.4 Типы парадигматических отношений

Устанавливают и четко разграничивают три типа парадигматических отношений между терминами (a), либо между понятиями (b) и (c):

a) отношения эквивалентности, которые применяются в случае как одноязычных, так и многоязычных ситуаций (см. разделы 8 и 9, соответственно);

b) иерархические отношения (см. раздел 10.2);

c) ассоциативные отношения (см. раздел 10.3).

Дальнейшее деление каждого из этих классов описано в разделах 8 — 10. Каждое из этих отношений должно быть охарактеризовано как симметричное (взаимное), и это должно быть обозначено с помощью системы меток, символов или сокращений, используемых для представления отношений в тезаурусе. В иллюстративных примерах ниже использованы метки, описанные в разделе 3.

5 Понятия и объем их значения в тезаурусе

5.1 Концептуальная основа

5.1.1 Главная прикладная задача тезауруса — это информационный поиск, целью которого является поиск понятий. Как было разъяснено в 4.1 и 4.2, понятия отображаются терминами. Каждый включенный в тезаурус термин должен отображать одно понятие (или единицу мысли). Понятия могут варьировать от простых (например, «кошки»), до очень сложных (например, «расовая дискриминация этнических меньшинств»). Для отображения более сложных понятий, как правило, требуются составные термины или фразы. Более полное обсуждение этих вопросов представлено в разделе 7.

5.1.2 Включенные в тезаурус понятия могут рассматриваться как принадлежащие к взаимоисключающим категориям, выявляемым на основе общих характеристик. Приводимый ниже набор примеров является иллюстративным и не является списком всех возможных категорий.

a) Предметы и их физические части:

Примеры —

1 birds	***птицы***
2 birth certificates	***свидетельства о рождении***
3 limbs	***части тела***
4 microforms	***микроформы***
5 monuments	***монументы***
6 mountain regions	***горные массивы***

Примечание — Физические части предметов также являются предметами и могут иметь свои собственные части.

b) Материалы:

Примеры —

1 adhesives	***клеи***
2 rubber	***резина***
3 titanium	***титан***

c) Действия и процессы:

Примеры —

1 dressmaking	***шитье одежды***
2 fertilization	***оплодотворение***
3 glaciation	***обледенение***
4 land management	***землеустройство***

d) События и явления:

Примеры —

1 birthdays	*дни рождения*
2 civil wars	*гражданские войны*
3 revolutions	*революции*

e) Свойства людей, предметов, материалов или действий:

Примеры —

1 consciousness	*сознательность*
2 elasticity	*эластичность*
3 personality	*индивидуальность*
4 speed	*скорость*

f) Дисциплины или предметные поля:

Примеры —

1 archaeology	*археология*
2 organic chemistry	*органическая химия*
3 theology	*богословие*

g) Единицы измерения:

Примеры —

1 hertz	*герц*
2 kilometers	*километр*

h) Типы людей и организаций:

Примеры —

1 charities	*благотворительные учреждения*
2 children	*дети*
3 international nongovernmental organizations	*международные неправительственные организации*
4 nations	*нации*
5 poets	*поэты*
6 visually impaired people	*лица с нарушениями зрения*

5.1.3 Уникальные сущности, обозначаемые с помощью имен собственных, также могут быть включены в тезаурус, при этом их следует использовать только для записей, в которых говорится о данном человеке, документе или предмете и т. п. Если обозначение некой сущности именем собственным не является темой (предметом) документа, а имеет к документу какое-то другое отношение, например, оно является автором (создателем) документа или входит в состав его названия, то такое имя собственное не должно стоять в одном ряду с предметными терминами, а должно быть приписано документу в другом месте.

Имена собственные также распределяются по категориям подобно понятиям в 5.1.2;

a) место.

Примеры —

1 Australia	*Австралия*
2 Milky Way	*Млечный Путь*
3 South Kensington	*Южный Кенсингтон*
4 Sri Lanka	*Шри Ланка*

b) специфические объекты, топографические характеристики и другие сущности.

Примеры —

1 Magna Carta	*Великая хартия вольностей*
2 Mona Lisa	*Мона Лиза*
3 Nelson's Column	*Нельсонова колонна*
4 Romeo and Juliet	*Ромео и Джульетта*
5 Skylab	*Скайлэб.*

с) физические лица, должности и юридические лица (организации),

Примеры —

1 Burns, Robert	*Бернс, Роберт*
2 United Nations Secretary-General	*Генеральный секретарь ООН*
3 World Health Organization	*Всемирная организация здравоохранения.*

5.2 Лексические примечания

Объем понятия, предусмотренный в тезаурусе, не всегда совпадает со значением, ассоциируемым в обычном употреблении с соответствующим дескриптором (или каким-либо из аскрипторов, заменяющих этот дескриптор). В тезаурусе контекст, как правило, понимаемый как иерархия вышестоящих и нижестоящих понятий, имеющих связи с рассматриваемым дескриптором, помогает уточнить желаемый объем понятия. Часто удается осуществить подбор такого однозначного дескриптора для определенного понятия, который доносит правильный выбор желаемого объема понятия до любого пользователя (см. 6.2.2). Однако в тех случаях, когда это сделать не удается, или когда дополнительная информация может внести ясность в вопрос о выборе определенного значения и сделать использование дескриптора более последовательным, целесообразно использовать поясняющие лексические примечания.

Лексические примечания следует использовать для уточнения границ понятия, особенно в тех случаях, когда значение дескриптора в обычном употреблении трактуется как более широкое, или более узкое, или когда требуется разграничить такие дескрипторы, значения которых в естественном языке имеют общую часть значения. Лексическое примечание также может быть использовано и для того, чтобы снабдить как индексатора, так и пользователя-поисковика рекомендациями, обеспечивающими использование одной и той же терминологии. Лексические примечания не должны давать полное определение, а призваны служить лишь разъяснением предполагаемого в тезаурусе использования термина. В тезаурусе отдельно могут быть представлены другие виды примечаний, а именно — определения (см. 6.2.3) и исторические справки (см. 6.2.4).

В отличие от квалификаторов, рассматриваемых в 6.2.2, лексические примечания не трактуются как составная часть того термина, к которому это примечание относится.

Примеры —

1 microwave frequencies *SN 1 GHz to 300 GHz*	*сверхвысокие частоты* *лп От 1 до 300 ГГц*
2 illuminations *SN Includes both the ornamental decoration and the illustrations in manuscripts, as well as in some early printed bucks, if done by hand*	*иллюминации* *лп Включают как декоративные украшения, так и иллюстрации в рукописях, а также в некоторых ранних печатных книгах, если выполнены вручную.*

5.3 Взаимные лексические примечания

Когда в примечании дается ссылка на другие понятия, то, как правило, каждое из этих понятий должно быть снабжено взаимным лексическим примечанием.

Примеры —

1 В английском языке:

food products

SN Use only for products intended for human consumption. For products for animals, see petfoods or feeds (Использовать только для продуктов, предназначенных для потребления человеком. Продукты для животных, см. petfoods или feeds).

petfoods

SN Food products for animals maintained as domestic pets. For products for human consumption, see food products. For products intended for non-domestic animals, see feeds (Продукты питания для животных, содержащихся в доме. Продукты для потребления человеком см. food products. Продукты, предназначенные для неодомашненных животных, см. feeds).

feeds

SN Products intended for non-domestic animals. For products for domestic pets, see petfoods. For products for human consumption, see food products (Продукты, предназначенные для неодомашненных животных. Продукты для домашних животных, см. petfoods. Продукты, потребляемые человеком, см. food products).

2 Во французском языке:

statistique

NE Désigne la science statistique; pour les données statistiques empoyer "statistiques" statistiques (Обозначает статистику как науку; для статистических данных использовать statistiques).

statistiques

NE Désigne les données statistiques; pour la science statistique, empoyer "statistique" (Обозначает статистические данные. Для статистики как науки использовать statistique).

3 В русском языке:

пища
лп Средства питания человека. Средства кормления животных, см. корм.
корм
лп Средства кормления животных. Средства питания человека, см. пища.

Даже если примечание требуется для уточнения только одного из понятий, все же полезно сделать редакционное примечание (см. 15.2.14) в записи каждого дополнительного понятия, где оно упомянуто. Взаимные ссылки употребляются для того, чтобы быть уверенным в том, что при изменении какого-либо понятия или при его удалении эти изменения отразятся и на других рассматриваемых понятиях.

6 Термины тезауруса

6.1 Форма термина

Термины, выбранные для представления понятий, могут быть как однословными, так и многословными (см. раздел 7).

6.2 Уточнение и устранение неоднозначности терминов тезауруса

6.2.1 Общие вопросы

В обычном языковом употреблении единичный термин может иметь более одного значения, и при этом самое востребованное общеязыковое значение не всегда соответствует тому понятию, которое востребовано в тезаурусе. Если контекст, обеспечиваемый иерархией вышестоящего и нижестоящего понятий, связанных с рассматриваемым понятием, недостаточен для выяснения предполагаемого объема понятия, то следует принять дополнительные меры. Поскольку все термины тезауруса должны быть выражены с наибольшей возможной степенью однозначности, особенно важно сформировать дескриптор для данного понятия таким образом, чтобы он делал понимание выбранного объема доступным для любого пользователя. Например, чтобы многозначный термин «депрессия» в зависимости от обстоятельств употребления можно было бы интерпретировать как «экономическая депрессия» или «метеорологическая депрессия». Для этого может быть использован квалификатор (см. 6.2.2). В тех случаях, когда это неудобно или недостаточно, или там, где дополнительная информация поможет разъяснить смысл и сделать использование более последовательным, следует сформулировать развернутое лексическое примечание (см. 5.2).

6.2.2 Омографы и реляторы (квалификаторы)

Омографы (иногда для них используют более широкий термин «омонимы») — это слова, имеющие одинаковое написание, но разные значения.

Пример — Cranes (Этот термин может относиться либо к птицам, либо к грузоподъемному оборудованию).

Если омографы используют в качестве терминов тезауруса, то значение каждого из терминов должно быть уточнено. При этом в качестве традиционного способа уточнения используется добавление заключенного в скобки релятора. Релятор должен быть как можно короче и в идеале состоять из одного слова. Часто, являясь более широким термином, релятор призван указывать на контекст или предметную область, к которой рассматриваемое понятие относится. Он не является лексическим примечанием, а входит в состав термина (см. п. 5.2).

Пример —

cranes (birds)	***лебедки (птицы)***
cranes (lifting equipment)	***лебедки (подъемные устройства).***

Релятор следует добавлять к каждому омографу, даже если в предметной области тезауруса один из его смыслов в большей степени привычен, чем все другие. Например, в английском языке «beams (structures)» — это дескриптор, который принят в инженерном тезаурусе, в который кроме того включен еще и термин «beams (radiation)». Например, во французском языке, «eau (boisson)» — дескриптор в экологическом словаре, в состав которого входит еще и дескриптор «eau (environnement)». В русском тезаурусе могут одновременно присутствовать дескрипторы «замок (крепость)» и «замок (запор)».

Если же один из терминов-омографов используется в специализированном тезаурусе, и его значение понятно пользователю тезауруса, релятор может быть опущен. Однако следует предусмотреть возможность того, что объем тезауруса в будущем может быть расширен, или что станет необходимым взаимодействие с другими словарями, охватывающими другие области.

При применении реляторов термины становятся несколько громоздкими, и поскольку некоторые электронные системы испытывают трудности в их применении, то следует избегать их (особенно в качестве дескрипторов) в тех случаях, когда можно найти другой способ устранения неоднозначности. По этой причине, использование многословного термина (как скоро такая многословная форма существует в естественном языке) предпочтительнее, чем использование однословного термина с релятором.

Пример — Термин «industrial plants» следует предпочитать термину «plants (facilities)». Последний термин, однако, может быть введен в качестве аскриптора. Аналогично русский термин «органы тела» предпочтительнее, чем «органы (анатомия)».

Заключенные в скобки реляторы не должны использоваться для образования инвертированных форм.

Пример — Неправильным является включение в тезаурус инвертированных форм «cookery (fish)» и «pens (fountain)». Вместо них следует использовать термины «fish cookery» (рыбные блюда) и «fountain pens» (авторучки). В последнем примере «fountain» используется для указания типа ручки, а не для устранения неоднозначности слова «pen». (См. также 7.7 о порядке слов в многословных терминах). Допустимым использованием реляторов с термином «pen» в английском языке являются термины «pens (enclosures)» — загоны и «pens (writing implements)» — средства письма. Правильное использование реляторов с термином «audition» во французском языке: «audition (physiologie)» и «audition (communication)». В русском языке: «корпус (тело)», «корпус (изделие)», «корпус (армейский)».

В случае сокращений и аббревиатур, которые часто являются омографами, релятором должна служить полная форма термина (см. также 6.6.8). Квалификаторы для акронимов и аббревиатур следует использовать только тогда, когда их значение пользователями тезауруса признается неоднозначным, т.е., когда у этих форм есть еще одно известное значение или в покрываемой тезаурусом области, или общем языковом употреблении.

6.2.3 Определения

Как правило, для уточнения того, как следует использовать дескриптор, не требуется наличия полного определения. Однако если по какой-то причине необходимо дать определение, то для него должно быть отведено отдельное поле, чтобы нельзя было спутать определение с лексическим примечанием. При каждом определении должен быть указан источник, из которого оно взято.

Пример —
Chiaroscuro
DEF The style of pictorial art in which only the light and shade are represented (OED)
кьяроскуро
Определение: Стиль живописи, использующий только свет и тень (OED).

Примечание — «OED» означает Oxford English Dictionary (gost.e-ecolog.ru ия подобная аббревиатура, встречающаяся в тезаурусном определении или лексическом примечании, должна быть объяснена во введении к тезаурусу (см. 13.4).

6.2.4 Исторические справки

Изменения терминов, возникающие во время обновления тезауруса, могут оказать влияние на эффективность поиска понятий. В таких случаях может помочь использование исторической справки.

Историческая справка должна относиться только к одному дескриптору, аскриптору или понятию. Ею следует пользоваться тогда, когда в тезаурус добавляется новый дескриптор или же существующий термин претерпел изменения, которые влияют на объем понятия в разные периоды использования тезауруса. Хотя подобную информацию можно включить в лексическое примечание, предпочтительнее создать историческую справку. Историческая справка может фиксировать дату введения термина или предоставлять более сложные рекомендации относительно того, как искать это же понятие на предшествующих и последующих этапах.

Примеры
1 Microwave ovens (микроволновые печи)
HN Concept introduced 1985 (Понятие введено в 1985 г.).
2 Notebook computers (ноутбуки)
HN Term introduced 1999; prior to that use «laptop computers» (Термин введен в 1999 г.; ранее использовался термин «laptop computers»).

6.3 Грамматические формы терминов

6.3.1 Существительные и именные словосочетания

Тезаурусный термин, как правило, представляет собой имя существительное или именное словосочетание. В частности он может быть выражен отглагольным существительным (герундием), (см. 6.3.4). В английском языке (и в других, таких как немецкий, французский и другие германские и романские языки)[1)] именные словосочетания встречаются в следующих двух формах:

a) В форме определительных словосочетаний (включая те, в которых в функции определения выступает существительное или притяжательная форма существительного[2)]).

Пример —

Burkitt's lymphoma	***лимфома Буркитта***
cold fusion	***холодная сварка***
environmental luminosity	***окружающая освещенность***
stone walls	***каменные стены***
tropical diseases	***тропические болезни.***

b) В форме предложных словосочетаний.

Пример —

accessories after the fact	***соучастники по факту (косвенные)***
hospitals for children	***больницы для детей.***

Предлоги придают термину излишнюю длину и неуклюжесть, поэтому по возможности следует избегать их использования, особенно в качестве дескрипторов. Например, термин «carbohydrate metabolism» следует предпочесть термину «metabolism of carbohydrates». (Впрочем, последний термин все еще мог бы быть принят в качестве аскриптора).

В некоторых других языках, например в китайском, предложные фразы отсутствуют, но имеются другие типы субстантивных (номинативных) словосочетаний, которые необходимо использовать в обсуждаемых случаях.

6.3.2 Прилагательные

Прилагательные используются в качестве компонентов именных словосочетаний, но при их использовании в качестве самостоятельного термина они могут стать причиной возникновения поисковых проблем. Так, поиск статьи по «использованию красных огней как сигналов предупреждения о низких мостах» с помощью таких терминов-прилагательных как «красный» и «низкий» может привести к тому, что мы получим информацию о низких сигналах и/или о красных мостах.

По этой причине использование прилагательных в качестве терминов тезауруса следует избегать. Такая же осторожность необходима и в применении не только прилагательных, но и адъективных существительных, например, «краснота» вместо «красный».

Исключения могут быть сделаны в случаях наличия веских причин, например, в случае использования тезауруса для индексирования отчетов о медицинских симптомах, коллекций изображений или других нетекстовых единиц, в которых прилагательные, описывающие внешний вид, могут играть очень важную роль.

Примеры —

1 Ddark red	***темно-красный.***
2 Large	***обширный.***
3 Oval	***овальный.***
4 Smooth	***гладкий.***

---

1) То же самое относится и к славянским языкам, включая русский.

2) Во всех этих языках наиболее типичны такие именные словосочетания, где в функции определения выступают прилагательные, но в этой функции могут использоваться и другие части речи, в частности существительные в косвенных падежах.

6.3.3 Наречия

Такие наречия как «очень» или «сильно» не должны использоваться в качестве терминов тезауруса. Словосочетания, начинающиеся с наречия, не допускаются в качестве терминов тезауруса, за исключением того случая, когда этому словосочетанию присваивается особое значение.

Примеры —
1 Very high frequency (сверхвысокие частоты).
2 Very large scale integration (интеграция сверхвысокого уровня).

6.3.4 Глаголы

В английском и русском языках глаголы, представленные в форме инфинитива или причастия, не должны использоваться в качестве отдельных терминов тезауруса. Однако допустимы термины, представленные в форме отглагольного существительного (например, «weaving», «broadcasting», «плетение», «вещание»). Действия следует обозначать существительными или отглагольными существительными. Для некоторых других языков может быть разрешен и инфинитив, но эти рекомендации должны быть сформулированы так, чтобы они не противоречили общепринятым соглашениям по индексированию.

Примеры

1 Cookery or cooking (не «cook», «to cook», «cooked», etc.)	*готовка пищи (не «готовить пищу»).*
2 Distillation (не «distil»)	*дистилляция (не «дистиллировать»).*
3 Swimming (не «swim»)	*плавание (не «плавать»).*

6.3.5 Артикли [1)]

6.3.5.1 Опущение артиклей

Как правило, следует избегать использования артиклей, предваряющих термины тезауруса. При необходимости следует использовать релятор, помещенный в скобки.

Примеры —
1 Arts, а не «the arts» (искусство).
2 State (political entity), а не «the state» (государство).

6.3.5.2 Сохранение артиклей

Если предваряющий артикль является неотъемлемой частью имени собственного и необходим в процессе поиска, то он должен быть включен в состав термина тезауруса в прямом порядке. В других случаях артикль должен быть опущен или термин должен быть представлен в инвертированном виде. Если опущение артикля приводит к неоднозначности, то следует добавить релятор. Считается ли артикль неотъемлемой частью имени, зависит от языка и от контекста. В следующих примерах языком тезауруса является английский.

Пример —
El Niño
Le Havre
Los Angeles
Needles (Isle of Wight)
The Who (rock music group)

Если при поиске нужно использовать термин как с артиклем, так и без него, должна быть сделана ссылка от аскриптора.

Пример —

Salvador (country)	*USE*	*El Salvador*
The Needles	*USE*	*Needles (Isle of Wight)*
Who, The	*USE*	*The Who (rock music group).*

6.4 Прописные буквы, знаки препинания и особые символы

6.4.1 Прописные буквы

В форме представления терминов тезауруса следует быть последовательным. Настоящий стандарт везде предусматривает нижний регистр, за исключением тех заглавных букв, которые необходимы

[1)] Правила употребления артиклей для русского языка не имеют значения, поскольку в русском языке нет артиклей.

в именах собственных[1]. Поскольку компьютерные технологии более не выдвигают требований по использованию только верхнего регистра, то следует использовать только нижний регистр с допущением первых заглавных букв.

Исключения могут быть сделаны для сокращений, аббревиатур, собственных имен или терминов, которые принято записывать с использованием особого стиля. В этих случаях должен быть использован тот стиль, который является наиболее распространенным среди предполагаемых пользователей тезауруса.

Примеры —
1 British Airways Plc.
2 ActiveX.
3 DNA.
4 NPK fertilizers.
5 pH.
6 Photocopies.

6.4.2 Неалфавитные символы

Использование знаков пунктуации, диакритических знаков и других специальных символов могут создают проблемы при вводе терминов, а также во время сортировки и обработки при проведении процедур поиска. Кавычки, скобки, дефисы и числовые символы усложняют реализацию процесса поиска. Использование таких символов должно быть минимизировано, особенно в дескрипторах.

Примеры — Замена специальных символов:

1 В английском языке:		
beta rays	***вместо***	***β-rays***
databases	***вместо***	***data-bases***
nonfiction	***вместо***	***non-fiction***
research and development	***вместо***	***research & development.***
2 Во французском языке:		
rayon bêta	***вместо***	***rayon β***
recherche et développement	***вместо***	***recherche & développement.***
3 В русском языке:		
бета-лучи	***вместо***	***β-лучи***
пятипроцентный раствор	***вместо***	***5%-ный раствор, или 5-процентный раствор***
стереоизображение	***вместо***	***3D-изображение.***

Однако, поскольку нельзя полностью избежать использования неалфавитных символов, они должны быть сохранены там, где терминология может быть двусмысленной, грамматически неправильной и неприемлемой с точки зрения сообщества пользователей тезауруса. Особенно в составе аббревиатур, химических названий, имен собственных, товарных знаков или в случае, если эти термины принадлежат к стандартизированным словарям данной сферы. Для тезаурусов на языках, отличных от английского, часто имеют очень большое значение диакритические знаки.

Примеры —
1 В английском языке:
2,4-D
bis(tributyltin) oxide
Boyle's law
Burkitt's lymphoma
X-rays.
2 Во французском языке:
2,4-D
oxyde de bis(tributylétain).
3 В русском языке:
п-ментила гидропероксид
N-метил-о-толуидин
метил(2-тиенил)дихлорсилан
ампер-час
Бозе — Эйнштейна статистика
критерий χ^2 .

[1] В немецком языке с заглавной буквы начинаются все существительные, которые в этой форме и включают в тезаурус.

Дефисы и знаки диакритики следует сохранять в аскрипторах, если они используются в составе общепризнанных орфографически правильных вариантов с точки зрения языка, используемого в тезаурусе.

Примеры — Использование специальных знаков в аскрипторах:

1 В английском языке:

non-fiction	***USE***	***nonfiction***
résumés	***USE***	***resumes.***

2 Во французском языке:

contre-marche	***EM***	***contremarche***
contre-plaqué	***EM***	***contreplaqué.***

3 В русском языке:

принцип д'Аламбера	***см.***	***принцип Даламбера***
2-бутанон	***см.***	***метилэтилкетон***
тимолфталексон S	***см.***	***метиловый синий.***

По возможности (но не в наименованиях химических веществ, а в других случаях, где они образуют неотъемлемую часть термина) скобки должны употребляться только в реляторах.

6.4.3 Иноязычные алфавиты

Если в тезаурусе для записи терминов применяют буквы различных алфавитов, то для каждого алфавита должен быть принят свой шрифт, позволяющий однозначно идентифицировать буквы, по начертанию сходные с буквами других алфавитов. В предисловии тезауруса необходимо указать, какой шрифт применяется для каждого алфавита. В русскоязычных тезаурусах для букв русского алфавита целесообразно использовать обычный шрифт, для латинских букв, например, — полужирный шрифт, а для греческих — полужирный с подчеркиванием.

Примеры —

1 Русские буквы:	*А*	*В*	*Е*	*К*	*М*	*О*	*Р*	*Т*
Латинские буквы:	***A***	***B***	***E***	***K***	***M***	***O***	***P***	***T***
Греческие буквы:	***Α***	***Β***	***Ε***	***Κ***	***Μ***	***Ο***	***Ρ***	***Τ***

2 Русские буквы: *ВВС* — ***военно-воздушные силы,*** *ВНР* — ***Венгерская народная республика.***

Латинские буквы: ***BBC — British Broadcasting Corporation, BHP — British horsepower.***

6.5 Единственное или множественное число

6.5.1 Культурные и лингвистические факторы

В естественных языках существуют разные правила, касающиеся использования единственного или множественного числа. В некоторых языковых сообществах, например во французском и немецком языках индексаторы, как правило, используют единственное число, для того чтобы пользователь мог использовать тезаурус так же, как он пользуется обычным словарем. Однако, в английском и испанском языках, выбор числа основывается на том, является ли тот или иной термин формой исчисляемого или неисчисляемого существительного. Принятие такого правила помогает отличать процесс, например «painting (рисование)», который может быть выражен только формой единственного числа, от результата этого процесса, в данном случае «paintings (рисунки)». В русскоязычных тезаурусах встречаются оба подхода.

Форма термина в каждом языке должна быть основана на принятых в данном языке правилах. Следствием такой практики является то, что в многоязычном тезаурусе термину в единственном числе на французском и немецком языках может соответствовать в других языках эквивалент в форме множественного числа.

Примеры —

1 fr:	***maison***
2 de:	***Haus***
3 en:	***houses***
4 es:	***casas***
5 ru:	***дома.***

6.5.2 Трактовка исчисляемых существительных

Примечание — Принципы, изложенные в настоящем подпункте, применимы при использовании русского, английского и испанского языков, а не французского и немецкого. Правила для других языков могут быть выработаны путем изучения практики, широко используемой в тезаурусах на этих языках.

Исчисляемые имена существительные — это наименования подлежащих счету объектов, к которым применим вопрос «How many? — Сколько?», а не «How much? — Как много?». Их следует обозначать формами множественного числа.

Примеры — Исчисляемые существительные:

1 documents	***документы***
2 penguins	***пингвины***
3 political parties	***политические партии***
4 windows	***окна.***

Исключением из этого правила являются наименования частей тела, которые обычно обозначаются формами единственного числа.

Примеры — Исключения:

1digestive system	***пищеварительная система***
2 eye	***глаз***
3 head	***голова.***

Другим исключением являются имена живых организмов. Многие виды, например, *Escherichia coli* или *Euonymus fortunei*, в обычном употреблении названия не имеют, и они известны только по их научным наимеваниям, для обозначения которых принято использовать латинский термин в единственном числе. Чтобы быть последовательными, в тех случаях, когда латинские наименования и названия, принятые в языке тезауруса сосуществуют в одном ряду, использование единственного числа может быть распространено на все объекты этого ряда. Однако использование единственного числа не является строго обязательным, и редакторы тезауруса могут отдать предпочтение множественному числу, например: «*zebras — зебры», «whales — киты»* или *«daffodils — нарцисс»*. Однако, как скоро соглашение принято, оно должно применяться последовательно.

Когда контрольный словарь используют для именования музейных предметов, как правило, применяют форму единственного числа терминов. В целом, однако, чаще понятия представляют не отдельные объекты, а категории объектов, и использование форм множественного числа является более целесообразным. Один предмет может быть назван «стул», но категорию скорее следует искать под обозначением «стулья». Использование форм множественного числа в дескрипторах в подобных случаях увеличит устойчивость принятых соглашений, поскольку оно позволяет использовать один и тот же тезаурус и для музейных предметов, и для других типов информационных ресурсов.

6.5.3 Трактовка неисчисляемых существительных

Примечание — Принципы, изложенные в настоящем подпункте, применимы для русского, английского и испанского языков, но не применимы для французского и немецкого языков. Для других языков наилучшие подходы могут быть выявлены путем изучения общей практики широкого использования тезаурусов на этих языках.

Неисчисляемые существительные — это наименования таких понятий, как материалы или вещества, которые отвечают на вопрос «How much? — Как много?», а не «How many? — Сколько?». Их, как правило, обозначают формами единственного числа.

Пример — Неисчисляемые существительные:

cotton	***хлопок***
aluminium	***алюминий***
steam	***пар.***

Однако если обслуживаемое тезаурусом сообщество пользователей рассматривает данное вещество или материал как класс с более чем одним элементом, то этот класс следует обозначать формой множественного числа.

Пример — Исключения:

grasses	***травы***
poisons	***яды***
steels	***стали.***

Наименования абстрактных понятий, свойств, систем убеждений, научных дисциплин, деятельностей и процессов часто являются неисчисляемыми существительными, и их следует обозначать формами единственного числа.

Примеры — Другие разряды неисчисляемых существительных:
1 Абстрактные понятия: personality; winter; индивидуальность; зима

2 Свойства: brittleness; opacity; solubility; хрупкость; непрозрачность; растворимость
3 Системы убеждений: Catholicism; Shintoism; communism; католицизм; синтоизм; коммунизм
4 Деятельности и процессы: cutting; immigration; shrinkage; отделение; иммиграция; сжатие
5 Научные дисциплины: astronomy; sociology; астрономия; социология.

Однако если рассматриваемое абстрактное понятие трактуется как класс с более чем одним членом, то этот класс следует обозначать формой множественного числа.

Пример — Дополнительные исключения:

chemical reactions	*химические реакции*
intelligence tests	*тесты интеллектуальности*
physical sciences	*физические науки.*

6.5.4 Сосуществование форм единственного и множественного числа

В любом языке, если формы единственного и множественного числа термина обозначают разные понятия, то обе эти формы должны входить в тезаурус. Различие между ними должно быть выявлено за счет добавления лексических примечаний и, если это возможно, квалифицирующего термина или фразы (релятора).

Пример —

В английском языке:
wood (material)
woods (areas of woodland)
Во французском языке:
statistique (science)
statistiques (données)
В русском языке:
лес (материал)
леса (территории).

Следует подчеркнуть, что добавленный релятор не является лексическим примечанием (см. 5.2), а становится неотъемлемой частью термина.

Если формы единственного и множественного числа термина обозначают одно и то же понятие и при этом их написание отличается таким образом, что в алфавитном списке их будут разделять не связанные с ними по значению термины, то следует ввести аскриптор, снабженный ссылкой.

Пример —

В английском языке:
mouse USE mice.
Во французском языке:
yeux EM oeil.
В русском языке:
уши см. ухо.

Если формы единственного и множественного числа обозначают одно и то же понятие и их написание отличается незначительно, обычно обеспечивается ввод только дескриптора. Но может быть добавлен и аскриптор, если предполагается компьютерное обращение к словарю.

6.6 Выбор дескриптора

6.6.1 Общие положения

Иногда в одном и том же языке одно и то же понятие может передаваться более чем одним термином. Если несколько терминов передают одно и то же понятие, то один из них должен быть признан дескриптором, а все другие (см. 8) — аскрипторами. Для многоязычного тезауруса эта ситуация описана в разделе 9.

Когда существует выбор между синонимическими формами выражения, должны учитываться предпочтения того сообщества, для обслуживания которого тезаурус создавался (чтобы исключить двусмысленности и выражения, которые пользователь может воспринять болезненно).

6.6.2 Правописание

Дескрипторами следует признавать наиболее распространенные варианты правильного написания слова. Если существуют другие варианты написания этого термина, которые являются общепризнанными, то их следует включать в тезаурус в качестве аскрипторов с указанием ссылок от аскрипторов к дескрипторам.

Примеры —

1 В английском языке:

Roumania	*USE*	*Romania*
Rumania	*USE*	*Romania*

Во французском языке:

Lithuanie	*EM*	*Lituanie*

В русском языке:

риелтор	*см.*	*риэлтор*

2 В английском языке:

non-fiction	*USE*	*nonfiction*

Во французском языке:

co-voiturage	*EM*	*covoiturage*

В русском языке:

Шри-Ланка	*см.*	*Шри Ланка.*

Правила орфографии должны соответствовать практике устоявшихся словарей и глоссариев языка тезауруса. Если между вариантами написания сделан выбор в пользу одного из диалектов (например, между американским и британским вариантами английского языка), то во введении тезауруса должно быть указано, какой из диалектов был выбран (см. 13.4). И это правило следует неукоснительно соблюдать.

Однако в именах собственных на законном основании могут допускаться непоследовательности, поскольку их написание должно быть приведено в соответствие с практикой признаваемой владельцами этих имен или нормативными источниками, такими как Virtual International Authority File (VIAF)[1)]. Примеры из набора 3, в которых принята британская орфографическая система за исключением имен собственных.

Пример — Британская орфографическая система за исключением имен собственных.

colour	*UF*	*color (цвет)*
defence	*UF*	*defense (оборона)*
Department of Defense	*(департамент правительства США)*	
Ministry of Defence	*(департамент правительства Соединенного Королевства).*	

Неправильное написание слов в дескрипторах недопустимо. Однако если орфографические ошибки распространены очень широко, то их включение в состав входов в тезаурус (особенно в электронный тезаурус) может оказаться целесообразным, и они могут быть введены в ранге аскрипторов.

Примеры —

1 В английском языке:

abattoirs	*UF*	*abatoirs* *abbatoirs* *abattoirs.*	*(скотобойни)*

2 Во французском языке:

Mitterrand François	*EP*	*Mitterand François (Франсуа Миттеран).*

3 В русском языке:

антена	*с*	*антенна*
Ельцин	*с*	*Ельцын.*

Термины с орфографическими ошибками могут служить входами в тезаурус только в следующих случаях:

a) орфографические ошибки достаточно часто встречаются;

b) не существует опасности установления связи между ошибочным написанием и термином, отличным от того дескриптора, с которым эта ошибка связана ссылкой в тезаурусе.

Такие входы должны быть помечены каким-то специальным образом, показывающим, что в данном случае мы имеем дело с орфографической ошибкой. В качестве одного из таких вариантов может быть замена метки UF меткой MS (означает «misspelling» — орфографическая ошибка), а в русском языке замена ссылки ***с*** (синоним) ссылкой ***оо*** (орфографическая ошибка).

[1)] VIAF является совместным проектом нескольких национальных библиотек [7]. VIAF включает более 13 миллионов официально зарегистрированных собственных имен, поддерживается консорциумом OCLC и доступен на сайте gost.e-ecolog.ru

Если неправильно написанные слова вводятся в качестве аскрипторов, то следует принять решение об исправлении таких орфографически неправильных аскрипторов в печатных и электронных версиях тезауруса, где они ничего не добавляют к пониманию пользователем термина, а использовать их только при поиске в Интернете для интерпретации запроса, предложенного пользователем.

6.6.3 Заимствованные термины и их перевод

Заимствованные термины, если они уже вошли в язык, могут использоваться в качестве дескрипторов (См. 9.3.3.2 и 9.3.3.3).

Примеры —

1 В английском языке:

bouquets
gestalt therapy
ombudsmen.

2 Во французском языке:

handball
permafrost
pizza.

3 В русском языке:

букет
омбудсмен
пицца.

Иногда заимствованный термин и предполагаемый перевод сосуществуют. Если заимствованный термин более широко распространен, он должен рассматриваться как дескриптор, однако и предлагаемый перевод может стать дескриптором, если он уже вошел в язык. Два таких термина следует связывать взаимными ссылками.

Примеры —

1 В английском языке:

coiffeurs	***USE***	***hairdressers***	***(парикмахеры)***
hairdressers	***UF***	***coiffeurs***	
abattoirs	***UF***	***slaughterhouses***	***(скотобойни)***
slaughterhouses	***USE***	***abattoirs.***	

2 Во французском языке:

baladodiffusion	***EP***	***podcast***	***(вещание через iPod)***
podcast	***EM***	***baladodiffusion.***	

3 В русском языке:

вертолет	***с***	***геликоптер***	
геликоптер	***см***	***вертолет.***	

6.6.4 Транслитерация

При транслитерации терминов из языков с другими алфавитами следует использовать утвержденные схемы соответствия алфавитных символов и их сочетаний. Во введении к тезаурусу должно быть указано, какая из утвержденных схем применяется в этом тезаурусе. Некоторые заслуживающие внимания схемы можно найти в справочнике [27] и в опубликованных стандартах, перечисленных в библиографии[1)]. См. также п. 12.4, где рассматриваются вопросы автоматического кодирования альтернативных письменностей.

6.6.5 Неологизмы, сленг и жаргон

Вместо быстросменяемых терминов, употребляемых в сленге или жаргоне, на практике может быть полезным использовать термины сленга/жаргона в качестве дескрипторов.

Однако часто в ответ на появление новых технологий или развития общественной и политической жизни появляются полезные неологизмы. Они могут быть приняты в качестве терминов тезауруса, если они будут оценены как полезные для поиска информации.

Примеры —

1 В английском языке:

carbon trading	***(торговля квотами на выброс углекислого газа)***
smart cards	***(смарт-карты).***

[1)] Для русского и других языков с кирилловским алфавитом действуют [8] и [9].

2 Во французском языке:
alicament
baladodiffusion (падкастинг, вещание через iPad)
littérisme.
3 В русском языке:
шестидесятники.

Просторечные и жаргонные формы могут быть применены в следующих обстоятельствах.

a) Вновь появившееся понятие выражается термином, возникшим в рамках определенной субкультуры или социальной группы, и не существует альтернативного общепринятого наименования для этого понятия. Тогда термину на сленге или жаргоне следует придать статус дескриптора.

Примеры —
1 Hippies (хиппи)
2 Web feeds (вэб-новости, вэб-каналы).

b) Если жаргонные термины широко используются вместо общепринятых терминов и многие пользователи считают полезным их использование в качестве входов в тезаурус. В этом случае принятому дескриптору следует обеспечить ссылку на аскриптор из сленга или жаргона.

Примеры —
1 В английском языке:

psychiatrists	*UF*	*shrinks*	*(психиатры)*
shrinks	*USE*	*psychiatrists.*	

2 Во французском языке:

policier	*EP*	*flic*	*(полицейский)*
flic	*EM*	*policier*	

3 В русском языке:

мобильник	*см*	*сотовый телефон*
сотовый телефон	*с*	*мобильник.*

c) Если целевая аудитория отдает явное предпочтение сленгу или жаргону, может быть полезным использование сленга/жаргона в качестве дескрипторов, интерпретируя альтернативные термины как аскрипторы. Такое решение может быть принято в случае, например, предпочитающей сленг молодежной аудитории.

6.6.6 Общепринятые названия и торговые марки

На практике оказывается, что продукт известен по торговой марке. Тогда в качестве дескриптора следует выбрать общепринятое название, а торговую марку следует включить в качестве аскриптора и только в том случае, если считается, что ее наличие облегчит пользователю доступ в тезаурус. Если зарегистрированные товарные знаки снабжены правовой защитой, чтобы избежать юридических проблем, к термину следует добавить символ «®».

Примеры —

1 soluble coffee	**UF**	***Nescafé®***	
2 Nescafé®	**USE**	**soluble coffee**	***(растворимый кофе).***
3 винчестер		**см**	**жесткий диск**
4 жесткий диск		**с**	***винчестер.***

Исключения допустимы, например, когда один универсальный продукт продается под несколькими конкурирующими торговыми наименованиями и пользователи хотят, чтобы эти разные бренды были разграничены. В этом случае дескриптором для универсального продукта может служить общее имя, а торговые наименования могут использоваться в качестве нижестоящих терминов (см. 8.4).

Пример —
ballpoint pens (шариковая ручка)
NT Bic pens
Biros.

6.6.7 Народные названия и научные наименования

Если как народное, так и научное название являются представителями одного и того же понятия, то предпочтение должно быть отдано той форме, в пользу которой высказываются пользователи тезауруса. Например, «penguins — пингвины» может быть выбран в качестве дескриптора в общем тезаурусе, а в зоологическом тезаурусе в качестве дескриптора предпочтительнее научный эквивалент, «Spheniscifоrmes». В таких случаях должны быть использованы взаимные отсылки.

6.6.8 Аббревиатуры и сокращения

Аббревиатуры и акронимы могут относиться не только к одному понятию и по этой причине бывают неоднозначными. Поэтому полная форма имени должна функционировать как дескриптор, а сокращенная форма — как аскриптор. При этом они должны быть связаны взаимными ссылками.

Пример — Акронимы/аббревиатуры не требующие разъяснения:

В английском языке:

United Nations Environment Programme
UF UNEP
UNEP
USE United Nations Environment Programme

direct current
UF DC (direct current)

DC (direct current)
USE direct current

national insurance
UF NI (national insurance)

NI (national insurance)
USE national insurance.

Во французском языке:

nouveaux pays industriels
EP NPI (nouveaux pays industriels)

NPI (nouveaux pays industriels)
EM nouveaux pays industriels.

В русском языке:

ЧП (чрезвычайное происшествие)
см чрезвычайное происшествие

чрезвычайное происшествие
с ЧП (чрезвычайное происшествие).

Когда сокращения или аббревиатуры широко известны, однозначны и легко понимаются всеми в пределах охватываемой тезаурусом области, особенно, если их употребление настолько утвердилось в языке, что полная форма термина употребляется очень редко или не употребляется совсем, тогда в виде исключения такие сокращения или аббревиатуры могут использоваться в качестве дескрипторов. Однако и в этом случае полная форма термина и его аббревиатура должны быть связаны взаимными ссылками.

Пример — Акронимы/аббревиатуры не требующие разъяснения:

В английском языке:

UNICEF
UF United Nations International Children's Emergency Fund
United Nations International Children's Emergency Fund
USE UNICEF

HIV
UF human immunodeficiency virus

human immunodeficiency virus
USE HIV.

Во французском языке:

SIDA
EP Syndrome d'immunodéficience acquise

Syndrome d'immunodéficience acquise
EM SIDA.

В русском языке:

ООН
с Организация Объединенных Наций

Организация Объединенных Наций
см ООН

СПИД
с синдром приобретенного иммунодефицита

синдром приобретенного иммунодефицита
см СПИД.

6.6.9 Имена собственные

6.6.9.1 Общие положения

Имена собственные часто бывают полезными в тезаурусе, но их подчас приходится исключать из-за того, что существует слишком много кандидатов, и особенно, если их формы контролируются другими авторитетными списками или наборами правил, подобных Anglo-American cataloguing rules (Англо-американские правила каталогизации)[1)] [10]. Имена собственные должны быть включены в тезаурус, если он является единственным средством для проверки точности терминов индексирования. Их включение имеет то дополнительное преимущество, что дает возможность установления иерархических или ассоциативных связей (см. раздел. 10) между предметными терминами и соответствующими собственными именами, например, когда последние являются отдельными представителями первых.

Примеры —

1	*mountains*		*горы*	
	NT	*Ben Nevis*	*н*	*Бен-Невис*
		Mount Everest		*Эверест*
2	*Hockney, David (1937-)*		*Хокни, Дэвид (1937 —)*	
	BT	*artists*	*в*	*художники*
	RT	*contemporary art*		*современное искусство*

6.6.9.2 Географические наименования

Названия стран и географических регионов нередко изменяются от языка к языку. Варианты терминов, обозначающих одни те же места, также могут применяться и в пределах одного языкового сообщества по следующим причинам:

a) в обиходном употреблении используются и «разговорный» и «официальный» варианты имени;

Пример —
Republic of Korea
South Korea.

b) исконная форма имени отличается от обычной формы имени, принятой в языке тезауруса;

Пример —
Leghorn (Английский тезаурус может включать этот перевод названия итальянского города, носящего местное название «Livorno».
Livorno (В тот же самый тезаурус может быть включено и исконное наименование).

c) в стране с двумя официальными языками могут сосуществовать две общепринятые формы.

Пример —
Gand (В Бельгии это французское название города Ghent сосуществует с голландским названием)
Gent (В Бельгии это голландское название города Ghent сосуществует с французским названием)

Статус дескриптора следует предоставить имени, которое наиболее привычно для большинства пользователей тезауруса. При этом предпочтение следует отдавать нормативным общеупотребительным, а не разговорным вариантам. Предпочтение должно быть отдано более короткому варианту нормативного общеупотребительного имени. За консультациями по поводу нормативных общеупотребительных вариантов имени следует обращаться к официальным источникам. Дескрипторы и аскрипторы должны быть связаны взаимными ссылками.

Пример —

Ireland		*Ирландия*	
UF	*Irish Republic*	*см*	*Ирландская Республика*
	Eire		*Эйре*

---

1) Этот набор правил широко использовался во время подготовки данной части настоящего стандарта, однако, существуют планы по его замене новым стандартом, известным как RDA: Resource Description and Access.

Eire	*Эйре*	
USE Ireland	*см*	*Ирландия*
Irish Republic	*Ирландская Республика*	
USE Ireland	*см*	*Ирландия.*

6.6.9.3 Имена юридических и физических лиц

Вариативность имен является распространенным явлением. В том случае, если она не взята под контроль, она может создавать трудности. При включении в тезаурус форма имен должна выбираться в соответствии с принятыми правилами каталогизации, например, таким как Англо-американские правила каталогизации (Anglo-American cataloguing rules [37]). Следует удостовериться в том, что имя, используемое в качестве дескриптора, является грамматически правильным, актуальным и достаточно полным, и что его нельзя перепутать с названиями других юридических и физических лиц. Общеупотребительные альтернативные формы имени должны получить статус аскриптора.

Примеры —		
1 Armstrong, Louis (1901-1971)	*UF*	*Satchmo (1901-1971)*
Армстронг, Луис (1901-1971)	*с*	*Сатчмо (1901-1971)*
2 Berners-Lee, Tim		
Бернерс-Ли, Тим		
3 Otlet, Paul (1868-1944)		
Отле, Поль (1868-1944)		
4 Centre national de la danse (France)	*UF*	*CND (Centre national de la danse)*
5 Association for Professional Broadcasting Education (U.S.)		*Association (U.S.) UF Broadcast Education*
6 Минобрнауки России	*см*	*Министерство образования и науки Российской Федерации.*

7 Сложные понятия

7.1 Общие положения

Понятия варьируются от очень простых до очень сложных. На полюсе простых понятий находятся такие понятия, как «серебро» или «люди». В терминах «грузинские серебряные чайники» или «защитники прав человека» представлены уже более сложные понятия. И в последнем из них три понятия объединены в одно более сложное понятие.

Сложные понятия очень часто передаются с помощью сложных терминов, которые могут быть морфологически расщеплены на два или более компонента. Одни из этих терминов состоят из нескольких слов, а другие из одного сложного слова.

«Биодеградация», например, — это однословный термин, в значении которого представление о приведении в негодность сочетается с представлением о том, что приведение в негодность обусловлено живыми организмами. Теоретически это понятие можно представить комбинацией понятий, образующих цепочку «приведение в негодность + живые организмы».

Разделение «биодеградация» на два понятия может показаться громоздким или неприемлемым, однако, когда некое сложное понятие постоянно передается с помощью многословного термина, выбор процедуры разделения на более простые понятия становится более легким и понятным. Так, например, понятие «защитники прав человека» можно легко разделить на понятия «права человека + защитники». А понятие «camping holilays» может быть адекватно выражено с помощью «camping + holilays». Для некоторых понятий существует много вариантов подобного синтеза. Например, понятие «управление человеческими ресурсами» может быть выражено с помощью «человеческие ресурсы + управление» или «люди + управление ресурсами» или даже «люди + ресурсы + управление».

При многократном выборе редактор тезауруса не сможет избежать принятия трудных и субъективных решений: следует ли применять комплексное понятие или же опираться на более простые понятия с тем, чтобы сами пользователи создавали их комбинации. Там, где тезаурус допускает много вариантов, решающим является желание сделать так, чтобы все пользователи для выражения данного сложного понятия применяли одну и ту же комбинацию. Быть последовательным становится еще труднее, если понятие может быть разбито более чем на два компонента, как например «управление человеческими ресурсами».

Еще одно соображение заключается в том, что введение сложных понятий, представленных как однословным, так и многословным термином, как правило, повышает специфичность. Большая специ-

фичность помогает пользователям достичь большей разрешающей способности. Например, если признать «biodegradability — подверженность биодеградации» более узким термином, чем «degradability — подверженность деградации», то релевантные документы становится искать проще, поскольку исключаются ненужные документы о приведении в негодность с помощью химических средств или из-за воздействия ультрафиолетовых лучей. В качестве платы за эту простоту объем тезауруса все увеличивается и увеличивается. Родственные понятия оказываются разделенными, что делает более трудным для пользователя процесс выбора правильного термина.

Большинство примеров в разделе 7 базируются на использовании многословных терминов. Однако те же самые соображения касаются и многих сложных понятий, передаваемых однословными терминами. Кроме того, в таких языках, как немецкий, для которого характерно наличие множества сложных слов, выбор разделения однословных терминов будет гораздо более широко представлен, чем в английском языке, и при этом будут работать те же самые руководящие принципы.

7.2 Природа составных терминов

В английском языке[1)] сложные термины состоят более чем из одного слова.

Примеры —

1 Road safety	***безопасность на дорогах***
2 Thesaurus management software	***программа ведения тезауруса***
3 Hirds of prey	***хищные птицы.***

Отдельные части большинства таких терминов вычленяются следующим образом.

a) Фокус или главное слово, это такой обозначенный существительным компонент, который определяет общий класс понятий, к которым этот термин может быть отнесен как единое целое.

Примеры —

1 Именной компонент «indexes — указатели» в термине «printed indexes — печатные указатели»

2 Существительное «hospitals — больницы» в предложном словосочетании «hospitals for children — больницы для детей».

b) Различительный признак или модификатор, т.е. один или несколько дополнительных компонентов, которые служат для сужения области фокусирования понятия, специфицируя тем самым один из его подклассов.

Примеры —

1 Прилагательное «printed — печатный» в термине «printed indexes» — печатные указатели

2 Сочетание предлога с существительным «for children — для детей» в термине «hospitals for children — больницы для детей».

Процедура подобного анализа иногда может быть применена к однословному термину, выраженному сложным словом.

Пример —

Термин	*Фокус*	*Дифференциальный признак*
agroindustry (агроиндустрия)	*industry (индустрия)*	*agro (префикс, происходящий из «agriculture»)*
biosecurity (биобезопасность)	*security (безопасность)*	*bio (префикс, происходящий из «biology»)*

7.3 Проблема допустимости использования сложных понятий

7.3.1 Обзор возможностей

В качестве простого примера рассмотрим ситуацию, когда такое понятие как «road safety (безопасность на дорогах)» могло бы быть представлено с помощью комбинации более простых понятий «road — дорога» и «safety — безопасность». Предполагается, что параллельно с увеличением сложности растет количество допустимых вариантов представления понятия. Так, понятие «thesaurus management software программа ведения тезауруса» может быть представлено как «thesauri — тезаурусы» + + «management — управление» + «software — программное обеспечение» или как «thesaurus management — ведение тезауруса» + «software — программное обеспечение».

[1)] Равно как и в других языках.

Необходимо рассмотреть следующие пять основных способов обращения с предлагаемыми сложными терминами.

a) Допустить использование сложных понятий, манипулируя этими образованиями как единым дескриптором.

Примеры —

1 Camping holidays	*отдых в кемпинге*
2 Road safety	*безопасность на дорогах*
3 Thesaurus management software	*программа ведения тезауруса.*

b) Допустить использование таких понятий, представляя их комбинацией двух или более терминов. В этом случае при вводе аскриптора для обозначения целостного понятия следует снабжать его указанием на соответствующую комбинацию более простых дескрипторов. Эту возможность часто трактуют как «расщепление» понятия.

Примеры —

1 Camping holidays	*USE*	*camping + holidays*
2 Thesaurus management software	*USE*	*thesaurus management + software*
3 Программы ведения тезауруса	*см*	*ведение тезауруса + программы.*

c) Допустить понятие в качестве кандидата с пересмотром этого решения по истечении испытательного срока.

d) Отклонить сложное понятие, если оно слишком редко употребляется, является нерелевантным или неуместным.

e) Исключить прямое включение понятия, но сохранить более широкое (вышестоящее) понятие, и разрешить использование аскриптора, связанного с вышестоящим дескриптором.

Примеры —

1 Road safety	*USE*	*safety*
2 Thesaurus management software	*USE*	*software*
3 Безопасность на дорогах	*см*	*безопасность*
4 Программы ведения тезауруса	*см*	*программы.*

7.3.2 Материал для размышления

Принятие разрешения об использования термина имеет сложный и субъективный характер. При принятии решения отчасти можно руководствоваться анализом файлов журнала поисковых транзакций, к которым обеспечен доступ (см. 13.1.3.3 и 13.2.2). Разработчик тезауруса должен соблюдать баланс преимуществ в поисковых ситуациях, возникновение которых можно предвидеть. При этом следует уделять внимание факторам, которые в порядке их приоритетности изложены ниже.

a) Частота использования. Первоочередному рассмотрению должны быть подвергнуты относительно сложные понятия с правами кандидата в дескрипторы в случае, если их часто ищут или если их значение для сообщества пользователей возрастает, особенно если термин, представляющий это понятие, широко используется и понятен аудитории. Обычно в таком случае используется вариант a) из 7.3.1, но также допустимы и варианты b) и c).

b) Степень точности. Если предстоит индексировать большой объем релевантных документов, то важным может оказаться решение о разграничении сложных понятий или набора близких понятий. Если в коллекции документов встречается термин «road safety (безопасность на дорогах)», то набор документов может затрагивать и понятия «rail safety (безопасность железнодорожного транспорта)» и/или более общее понятие «transport safety (транспортная безопасность)», а также и понятия «fire safety (пожарная безопасность)» и «safety in the home (безопасность в быту)». Возможно, многие пользователи захотят уметь различать эти разные аспекты безопасности. За счет включения в тезаурус всех этих терминов достигается более высокий уровень специфичности (конкретности), позволяющий проводить как индексирование, так и поиск с большей точностью.

c) Неполнота поиска как результат множественности вариантов поиска. Обеспечение доступа к набору близкородственных дескрипторов затрудняет выбор вариантов. Например, если в тезаурусе присутствуют одновременно все термины: «safety», «transport safety», «passenger safety», «pedestrian safety», «vehicle safety», а также «road safety» («безопасность», «безопасность на транспорте», «безопасность пассажиров», «безопасность пешеходов», «безопасность автомобиля», а также «безопасность дорожного движения»), то, чтобы быть уверенным, что найдены все нужные материалы, при поиске придется проверить все эти варианты (а также комбинации термина «safety» с такими терминами, как «roads», «vehicles», «cars», «road traffic»).

Если общее количество документов по безопасности (safety) невелико, то может оказаться, что поиск был бы проще и эффективнее, если бы ни один из сложных терминов не был допущен в качестве дескриптора. Тогда единственной поисковой возможностью был бы выбор посткоординированного поискового предписания «roads + safety» («дороги + безопасность)».

d) Сложные термины, в которых один фокус характеризуется более чем по одному дифференциальному признаку. В предыдущих примерах были использованы такие относительно простые термины, в которых фокус «безопасность» квалифицировался только по одному дифференциальному признаку, например «транспорт». Большая сложность возникает с такими понятиями, как «транспортное средство пожарной безопасности», в котором основное значение (фокус) квалифицируется одновременно по двум дифференциальным признакам. Использование таких терминов ведет к дальнейшему умножению количества вариантов и мешает проведению простого поиска, поэтому их следует избегать. Например, сочетание терминов «транспортное средство» + «пожарная безопасность» адекватно отражает необходимое понятие и не создает путаницы.

e) Ожидания предполагаемых пользователей. Такой термин, как «pre-coordinate indexing — предкоординатное индексирование» мог бы быть приемлемым и полезным в тезаурусе, ориентированном на специалистов в области информационного поиска, но если тезаурус ориентирован на более широкую аудиторию, использование этого термина может вызвать недоумение. В последнем случае лучше не разделять этот термин на термины «pre-coordination — предкоординация» + «indexing — индексирование», а выбрать вариант c), создав входные термины к более широкому термину «индексирование» (см. 8.4).

f) Предотвращение неоднозначности терминов. Иногда включение сложного понятия необходимо для того, чтобы избежать неоднозначности, возникающей в том случае, если какой-либо компонент этих понятий допускает различные способы комбинирования при употреблении в своих разных значениях. Например, комбинация терминов «libraries (библиотеки)» и «science (наука)» могут использоваться для формирования и термина «library science (библиотечная наука)», и термина «science libraries (научные библиотеки)». Во избежание получения в качестве результата поиска ненужных фрагментов и поискового шума, одно или оба эти сложные понятия могут быть признаны дескрипторами.

g) Осложнения, связанные с многоязычностью (см. раздел 9). В случае многоязычных тезаурусов, присущие какому-либо из языков ограничения или особые потребности могут оказывать влияние на включение или исключение сложного понятия во всех языках тезауруса.

h) Ожидаемая поисковая среда. Если программное обеспечение поиска предлагает сложные функции, такие как кластеризация результатов поиска, адаптация к интересам отдельного пользователя или сообщества пользователей это может повлиять на уровень необходимой для тезауруса специфичности (конкретности). Этот фактор должен быть уравновешен допущением того, что программное обеспечение и другие элементы среды в будущем могут измениться.

7.3.3 Обстоятельства, которые способствуют расщеплению сложных понятий

Сложные понятия подлежат разделению, если складываются следующие условия:

a) разделение допускается в случае весьма специфичных понятий, не принадлежащих к центральной части, сердцевине предметного охвата тезауруса. Однако включение в тезаурус большого числа таких периферийных терминов увеличивает общий объем и сложность словаря, не внося существенного улучшения поисковых характеристик;

b) если предполагается, что предлагаемый термин годится для индексирования лишь весьма незначительного количества документов, то вряд ли стоит присваивать этому термину статус дескриптора, а следует реализовать вариант разделение понятия;

c) сложное понятие следует разделить, если фокус понятия охарактеризован с помощью более чем одного дифференциального признака, положенного в основу классификации понятий. Например, термин «underwater cine cameras» должен быть разделен на термины «underwater cameras» и «cine cameras»;

d) комплексное понятие принято разделить, если фокус этого понятия является свойством, частью или компонентом дифференциального признака.

Примеры —

1 Aircraft engines	***двигатели самолетов.***
2 Hospital floors	***больничные отделения.***
3 Instrument reliability	***надежность инструмента.***
4 Soil acidity	***кислотность почвы.***

Допускаются и исключения для случаев, когда мы имеем дело с понятием, своеобразие которого выявляется из четкого определения этого понятия, например, как в случае понятия «температура тела». Очевидно, что ссылка на температуру тела в корне отличается от ссылки на температуру других неодушевленных предметов или веществ, обсуждаемых в том же документе данной коллекции.

7.3.4 Обстоятельства, которые препятствуют расщеплению сложного понятия

Сложные понятия не подлежат разделению, если складываются следующие условия.

a) Термин, служащий для выражения какого-либо понятия привычен в повседневном использовании или в области, охватываемой тезаурусом, так что дробление этого термина затруднило бы процесс понимания.

Примеры —

1 Data processing	***обработка данных***
2 Gross domestic product	***внутренний валовой продукт.***

b) Разделение понятия на составляющие привело бы к утрате смысла, или к двусмысленности.

Примеры —

1 В английском языке:

plant food — растительная пища
(«food» + «plants» может означать и «растения как еда», и «продукты питания для фабрик»).

2 Во французском языке:

marque de voiture — марки автомобилей
(«marque» + «voiture» может означать и «марка автомобиля», и «автомобиль как марка престижа»).

société de construction
(«société» + «construction» может означать и «société» для «construction» — строительная фирма, и «construction» для «société» — устройство общества).

3 В русском языке:

пища животных
(«пища» + «животные» может означать и «пища» для «животных», и «животные» для «пищи»).

c) Понятие обозначается дескриптором, являющимся или именем собственным, или включает в свой состав имя собственное.

Примеры —

1 Boolean logic	***булева логика***
2 United Nations	***Организация Объединенных Наций***

d) Дифференциальный признак в рассматриваемом термине утрачивает свой первоначальный смысл.

Примеры —

1 В английском языке:

lawn tennis	***теннис, не «газонный теннис»***
deck chairs	***шезлонги, не «палубные кресла»***
trade winds	***пассаты, не «торговые ветра».***

2 Во французском языке:

pistolet à température	***бесконтактный измеритель температуры***
maison mère	***головная компания, не «материнский дом»***
police d'assurance	***страховой полис, не «страховая полиция».***

3 В русском языке:

черный ящик	***(система со скрытой структурой)***
Большой взрыв	***(гипотеза происхождения Вселенной)***
белый шум	***(шум с равномерным спектральным распределением).***

e) Понятие представлено термином, характеризующий дифференциальный признак которого построен на метафорическом подобии с некоторым посторонним предметом или событием.

Примеры —

1 Tree structures (древовидные структуры).

2 Wing nuts (гайки-барашки).

f) Если слова, входящие в состав имеющего терминологическое употребление словосочетания, рассматривать по отдельности, обозначают такие понятия, которые не имеют никакой связи с понятием, обозначенным словосочетанием, включающим их в свой состав. В этих случаях было бы неверным индексировать документы, содержащие термин-словосочетание, с помощью подобных терминов-частей.

Примеры —

1 В английском языке:

fire escapes — пожарные лестницы (обсуждение пожарных лестниц может ничего не говорить о «fires — огонь, пожары» или об «escaping — побег, бегство»)

swimming pools — плавательные бассейны (обсуждение плавательных бассейнов может содержать мало информации о плавании).

2 Во французском языке:

établissement de santé — учреждение здравоохранения (обсуждение учреждений здравоохранения может содержать мало информации о «santé — здоровье»)

classe de neige (обсуждение «classe de neige» может содержать мало информации о «neige — снег»).

3 В русском языке:

лазерный диск (обсуждение лазерных дисков может содержать мало информации о лазерах)

настольная лампа (обсуждение настольных ламп не имеет отношения к сведениям о столах).

g) Понятие представлено термином, фокус которого имеет два разных смысла в случаях присутствия и отсутствия дифференцирующего признака.

Пример —

artificial flowers (искусственные цветы)	***— искусственные цветы не являются цветами, но, например, в тезаурусе по дизайну интерьера это понятие вполне может быть подведено под класс «цветы (элемент декора)».***
chocolate eggs (шоколадные яйца)	***— было бы ошибочным использовать термин «яйца» для индексирования текстов, в которых идет речь о шоколадных яйцах, т. к. эти «яйца» — вид кондитерских изделий, а не яиц).***

7.4 Способы расщепления сложных понятий

В некоторых простых случаях, например, в случае термина «road safety», обозначенное с помощью этого многословного термина понятие может быть адекватно выражено и с помощью комбинации отдельных входящих в состав этого термина слов «roads» + «safety». Обоснованием для того, чтобы поступать в рамках отдельного тезауруса подобным образом, может служить то, что любой документ, касающийся «road safety», содержит сведения, релевантные при поиске сведений, касающихся как термина «road», так и термина «safety». Человек, ищущий информацию о дорогах, сочтет релевантной для своего запроса статью о безопасности на дорогах.

Если понятие разделено, то его составляющие могут быть обозначены такими дескрипторами, которые не полностью совпадают со словами, входящими в состав расщепляемого многословного термина. Чтобы приспособить слова, входящие в состав многословного термина, для обозначения нужных понятий, их приходится модифицировать. Например, «rail safety — безопасность на железных дорогах», не должно быть разделено на «rails — рельсы» + «safety — безопасность», но должно быть выражено как «railways — железные дороги» + «safety — безопасность». Для термина «human resources — человеческие ресурсы» комбинация «people — люди» + «resources — ресурсы» вероятно более приемлема чем «humans — человеческий» + «resources — ресурсы».

Для сложных понятий, которые не заслуживают присвоения им статуса единого дескриптора, но при этом пользователи тезауруса, скорее всего, будут заинтересованы в их поиске, следует предусмотреть в тезаурусе записи следующего вида (см. раздел 8.5).

Пример —

rail safety		***безопасность на железных дорогах***
USE railways	***см***	***железные дороги***
+ safety		***+ безопасность***

7.5 Сохранение компонентов понятия

7.5.1 Общие вопросы

В случае если составной термин принят в качестве дескриптора, следует рассмотреть вопрос о целесообразности включения в тезаурус в качестве дескрипторов и компонентов этого термина, если они еще не были ранее включены в тезаурус. Например, если мы принимаем дескриптор «road safety — безопасность на дорогах», то в тезаурус обычно включают дескрипторы «roads — дороги» и «safety — безопасность». При этом последний термин следует признать вышестоящим дескриптором по отношению к дескриптору «road safety — безопасность на дорогах», а дескриптор «roads — дороги» ассоциативно связанным с этим сложным термином (см. 10.3).

7.5.2 Части и компоненты

Путаница может возникнуть, когда в качестве дифференциального признака составного термина использует название машины или какого-либо другого сложного агрегата, а фокус термина обозначает составную часть. Например, если «aircraft engines — двигатели» разделяется на «aircraft — самолет» и «engines — двигатели», при поиске по «aircraft — самолет», рассматриваемом как отдельный термин, выданы будут записи обо всех деталях воздушных судов наряду с документами о воздушных судах в целом.

Для некоторых собраний документов это может оказаться приемлемым, но для других это приведет к перегрузке термина «самолет», и исчезнет возможность изолированного рассмотрения документов, в которых воздушные суда рассматриваются как единое целое. Одним из возможных выходов из такой ситуации является создание дескриптора «aircraft components — компоненты самолетов», который можно использовать в сочетании с дескрипторами «engines — двигатели», «instrumentation — приборы», «under-carriages — шасси» или названиями любого другого компонента воздушного судна.

7.6 Последовательность в трактовке сложных понятий

Последовательность в разработке тезауруса всегда полезна и ведет к последовательности в практике применения. Однако, в случае установления целесообразности включения в тезаурус сложных понятий полной последовательности достичь трудно, и она не всегда необходима. Присутствие непоследовательностей неизбежно, поскольку центральным областям тезауруса специфичность требуется в большей степени, чем его периферийным областям. Такие непоследовательности не оказывают негативного влияния на процесс информационного поиска, и ими можно пренебречь, если имеется достаточно прозрачных и полезных входов в тезаурус.

Последовательные правила оказываются полезными при редактировании тезауруса. Например, отдельному человеку трудно судить о необходимом уровне специфичности (конкретности), когда тезаурус является общим для нескольких организаций, которые используют его при работе с отличающимися друг от друга коллекциями документов и с разными группами пользователей. Аналогичным образом, наличие общего руководства необходимо, если составление тезаурусов является сетевым проектом, в котором задействованы сотни участников из разных стран.

Для достижения согласованности следует создавать набор критериев, соответствующих предметной области, примером которого может служить проект Art & Architecture Thesaurus [26], где представлен набор правил разделения составных терминов с особым вниманием к потребностям каталогизации единиц хранения в музеях и галереях. Одним из таких правил является разделение составного термина, если различительным признаком является стиль или название периода создания, вследствие которого понятие «барочная позолота» будет представлено как сочетание «барокко» + «позолота».

7.7 Порядок слов в многословных терминах

В дескрипторах, являющихся либо адъективными, либо предложными именными словосочетаниями, должен быть сохранен естественный порядок слов. Термины не должны подвергаться инвертированию. Инвертированная форма предложных словосочетаний может быть включена в тезаурус, но в качестве аскриптора. Это особенно полезно, если тезаурус используется в печатной форме, но может не потребоваться для электронного тезауруса, в котором возможен поиск по любому составному элементу входящего в состав тезауруса термина.

Примеры —

1 Matter, states of	***USE***	***states of matter***
дела, состояние	***см***	***состояние дел***

2 Prey, birds of	*USE*	*birds of prey*
птицы, хищные	*см*	*хищные птицы.*

В адъективных словосочетаниях (например, «brown bread — черный хлеб» или «double beds — двуспальные кровати») инвертирование вряд ли необходимо, поскольку существительное, являющееся фокусом адъективных словосочетаний, обычно входит в состав тезауруса в качестве вышестоящего дескриптора (см. 10.2). Статья вышестоящего термина (например, «bread — хлеб» или «beds — кровати») является входом, по которому могут быть найдены все нижестоящие термины.

8 Отношения эквивалентности в одноязычном контексте

8.1 Общие положения

Отношения эквивалентности — это отношения между дескриптором и соответствующим ему аскриптором(ами) в пределах одного и того же естественного языка. Как продемонстрировано в модели данных в разделе 15, это скорее отношения между терминами, чем отношения между понятиями.

Примечание — Отношения эквивалентности, которые связывают между собой термины из разных естественных языков, которые обозначают одно и то же понятие и которые часто получают название «межъязыковая эквивалентность», описаны в разделе 9.

Отношение эквивалентности, устанавливаемое между дескрипторами и аскрипторами, является взаимообратным отношением. И эту взаимную (обратную) связь обозначают ссылками, приведенными ниже.

USE предшествует дескриптору;

UF (use for *или* used for) предшествует аскриптору[1)].

Пример —

В английском языке:

greenhouses		***(теплицы)***
UF	***glasshouses***	

glasshouses	
USE	***greenhouses***

Во французском языке:

biodiversité		***(биоразнообразие)***
EP	***diversité biologique***	

diversité biologique	
EM	***biodiversité***

В русском языке:

парники	
с	***теплицы***

теплицы	
см	***парники.***

Эквивалентность устанавливается в следующих случаях:

a) термины являются синонимами (см. 8.2);

b) термины являются квази-синонимами (см. 8.3);

c) термин считается излишне специфичным (детализированным) и его представляют с помощью другого термина с более широким значением (см. п. 8.4);

d) термин считается излишне специфичным и представляется с помощью комбинации двух или более терминов (этот случай известен под названием «compound equivalence — составная эквивалентность») (см. 8.5).

8.2 Синонимы

В практической работе встречаются синонимы разных типов. Следующий ниже список указывает наиболее типичные виды синонимов.

[1)] Здесь указаны обозначения связи аскриптора с дескриптором только для англоязычных тезаурусов. Соответствующие обозначения для других языков см. в таблице 2 настоящего стандарта, а также в примерах данного пункта.

a) Термины, происходящие из разных языков.

Пример —

В английском языке:

freedom; liberty (свобода)
sweat; perspiration (пот).

Во французском языке:

marketing; mercatique (маркетинг)
lobby; groupe de pression (лобби).

В русском языке

менеджмент, организационное управление
лингвистика, языкознание.

b) Общепринятые и научные названия.

Пример —

В английском языке:

baking soda; sodium bicarbonate (сода пищевая)
rock roses; Cistus (каменная роза, ладанник).

Во французском языке:

aspirine; acide acetylsalicylique (аспирин)
ail; Allium sativum (чеснок).

В русском языке:

пищевая сода, натрия гидрокарбонат
аспирин, ацетилсалициловая кислота.

c) Общепринятые названия и торговые марки.

Пример —
vacuum flasks; Thermos® flasks; Thermos® bottles
жесткий диск, Винчестер.

d) Варианты названия для вновь возникающих понятий.

Пример —

hovercraft; air cushion vehicles
laptop computers; notebook computers
ховеркрафт; судно на воздушной подушке
переносные компьютеры; наколенные компьютеры; ноутбуки.

e) Недавно возникшие или предпочтительные термины в сравнении с устаревшими или вышедшими из употребления терминами.

Пример —

developing countries; underdeveloped countries
radio; wireless
развивающиеся страны; неразвитые страны
чахотка; туберкулез.

f) Варианты написания: в том числе варианты написания основ слов, инверсия порядка слов и неправильные формы множественного числа. Возможно, полезным было бы включение в число вариантов написания также некоторых типичных орфографических ошибок (см. 6.6.2).

Пример —

В английском языке:

geese; goose (гуси; гусь)
groundwater; ground-water; ground water (грунтовые воды)
paediatrics; pediatrics (педиатрия)
radiation, ionizing; ionizing radiation; ionising radiation (ионизирующая радиация)
Romania; Rumania; Roumania (Румыния)

Во французском языке:

clef; clé (ключ)

Lithuanie; Lituanie *(Литва)*
nénuphar, nénufar *(кувшинка)*
bioénergie, bio-énergie *(биоэнергия)*
oeil, yeux *(глаз; глаза)*

В русском языке:
ухо; уши
искусство; искуство
Санкт-Петербург; Петербург
д'Аламбер; Даламбер
юго-восток; юговосток.

g) Термины, восходящие к различным культурным традициям, которые обслуживаются общим языком.

Пример —

В английском языке:
flats; apartments *(квартиры)*
lifts; elevators *(лифты).*

Во французском языке:
mitaine (fr-CA); moufle (fr-FR) *(рукавица)*
journal de classe (fr-BE); agenda scolaire (fr-FR) *(школьный дневник).*

В русском языке:
изба; хата
подъезд; парадная
алгоритм; алгорифм.

h) Аббревиатуры и акронимы или их полные наименования.

Пример —

FAO; Food and Agriculture Organization
pvc; polyvinyl chloride
ПВХ; поливинилхлорид
ООН; Организация Объединенных Наций
вуз; высшее учебное заведение.

i) Общеупотребительные названия или сленг и жаргон.

Пример —

В английском языке:
psychiatrists; shrinks *(психиатры)*
soluble coffee; instant coffee *(растворимый кофе).*

Во французском языке:
policier; flic *(полицейские)*
café soluble; café instantanné *(растворимый кофе).*

В русском языке:
следователь; сыщик; детектив
гибкий магнитный диск; дискета.

В этих случаях выбор дескрипторов определяется потребностями подавляющего большинства пользователей и проводится с учетом рекомендаций, изложенных в 6.4 и 6.6. Чтобы обеспечить предсказуемость, эти критерии должны применяться последовательно на всех этапах создания тезауруса. Если было принято решение, что в качестве дескрипторов следует выбирать общеупотребительные, а не научные названия, то такой выбор должен проводиться последовательно за исключением лишь тех случаев, когда соответствующих общеупотребительных названий в языке просто не существует. Аналогичным образом, должны быть приняты и последовательно проведены соглашения, касающиеся орфографии. Соглашения и критерии, касающиеся редакторских правил, должны быть приведены во введении к тезаурусу (см. 13.4). В качестве меток, отличных от UF (см), допускаются указания на то, какой тип эквивалентности превалирует в отношениях между членами пары терминов. А именно, в качестве таких меток иногда используются (в английском языке) следующие символы:
SP (SPelling variant) — вариант написания,
MS (MisSpelling) — орфографическая ошибка (см. 6.6.2),

AB (Abbreviation) — сокращение,
FT (Full form of the Term) — полная форма термина.

Для того чтобы распространить этот принцип и на все случаи, описанные в пунктах 8.3 и 8.4, необходимо добавить описание дополнительных меток. Однако, адаптация должна проводиться лишь с очень большой осторожностью (см. 10.4). Во введение к тезаурусу должна быть помещена информация о нестандартных метках (см. 13.4).

8.3 Квазисинонимы

Квазисинонимы часто занимают разные точки континуума значений и могут оказаться даже антонимами.

Примеры — Противоположности, трактуемые как квазисинонимы:
1 consistency, inconsistency
wetness, dryness.
2 состоятельность, несостоятельность
сырость, сухость.

Во второй строке приводимого выше примера мы находим тезаурусное понятие «количество влаги в объекте или материале», которое может быть выражено в терминах, говорящих либо о влажности, либо о сухости. Выбор обозначения одного из этих аспектов в качестве дескриптора, а другого в качестве аскриптора носит произвольный характер.

Но существуют и такие случаи, когда понятия настолько тесно связаны, что представляется очевидным, что обсуждение одного из этих двух понятий заинтересует и пользователей, занятых поиском парного понятия.

Пример — Близкие понятия, трактуемые как квазисинонимы:

bushes, shrubs	***кусты, заросли***
mallets, hammers	***молотки, молоты***

Диапазон отождествления квазисинонимов в некоторой степени зависит от предметной области тезауруса. В тезаурусе, специализирующем на швейном производстве, термины «перчатки» и «варежки» следует признать отдельными дескрипторами. В то же время в более широком тезаурусе текстильной отрасли эти два термина могут трактоваться как квазисинонимы. Принятие решения должно основываться на желаемой степени точности поиска.

8.4 Включение специальной терминологии в состав вышестоящего понятия

Иногда целесообразно объединять в рамках множества эквивалентов как название класса, так и названия членов этого класса, выбрав при этом в качестве дескриптора более широкий термин, стоящий в иерархии выше.

Пример —

rock			***камень***		
UF	***basalt***		***с***	***базальт***	
	granite			***гранит***	
	slate			***сланец***	
	etc.			***и т. д.***	
basalt	***USE***	***rock***	***базальт***	***см***	***камень***
granite	***USE***	***rock***	***гранит***	***см***	***камень***
slate	***USE***	***rockрок***	***сланец***	***см***	***камень***

Такой технический прием иногда используют для уменьшения количества дескрипторов в тезаурусе. Его применение ограничивает точность, которая может быть достигнута в процессах индексирования и поиска, и обеспечивает преимущество лишь в том случае, когда в индексируемой коллекции содержится очень мало информации по рассматриваемой тематике. Наличие аскрипторов создает дополнительные точки входа в систему, чтобы обеспечить пользователю доступ к нужной информации. Однако, подобное «поглощение» нижестоящих понятий вышестоящими не следует использовать при обилии информации по заданной теме; необходимость достижения большой степени точности поиска требует аккуратности указания понятий.

8.5 Представление сложных понятий с помощью сочетания терминов

Если многословный термин оказывается непригодным для роли дескриптора, и в то же время может быть востребован некоторыми пользователями, то его разрешается представить в виде комбина-

ции двух или более дескрипторов (это называют составной эквивалентностью [см. 7.3.1, перечисление b)]. В тезаурус могут быть введены словарные статьи следующего типа:

Пример — Представление сложных понятий:

coal mining		***добыча угля***	
USE	***coal***	***см***	***уголь***
+	***mining***	***+***	***добыча полезных ископаемых***
ferromagnetic films			***ферромагнитная пленка***
USE	***ferromagnetic materials***	***см***	***ферромагнитные материалы***
+	***films***	***+***	***пленка.***

При этом следует предусмотреть обратные статьи следующего типа:

Пример — Представление компонентов понятий:

coal		***уголь***	
UF+	***coal mining***	***см+***	***добыча угля***
mining		***добыча полезных ископаемых***	
UF+	***coal mining***	***см+***	***добыча угля.***

Трехчленные отношения признаются сложными, а разграничение меток «**USE**» и «**+**» введено только для облегчения чтения. В функциональном плане отношения между «coal» и «*coal mining*» идентичны отношениям между «mining» и «*coal mining*» в обоих направлениях.

Гипотетически некоторые термины могли бы быть представлены в тезаурусе комбинаций дескрипторов, связанных отношением ИЛИ, а не И (см. примеры ниже). Но эту ситуацию следует рассматривать как случай омографии. Связей ИЛИ в комбинациях дескрипторов не следует допускать, вместо этого следует формировать тезаурусные статьи, в которых будет указан выбор определенной альтернативы значения.

Пример — Практика, которую следует избегать:

В английском языке:		
pitch		
USE	***audio frequency***	***(частота звука)***
OR	***gradient***	***(наклон)***
Во французском языке:		
ton		
EM	***nuance de couleur***	***(оттенок цвета)***
OU	***tonalité (musique)***	***(тональность)***
В русском языке:		
тон		
см	***оттенок цвета***	
или	***музыкальный звук.***	

Однако этот тип комбинирования понятий следует избегать, и вместо этого вводить пояснения, которые определяют выбор одного из альтернативных значений термина.

Пример — Разъяснение альтернативного значения термина:

В английском языке:			
	pitch (sound)	***высота (звук)***	
	USE audio frequency	***см***	***частота звука***
	pitch (steepness)	***крутизна (наклон)***	
	USE gradient	***см***	***уклон***
Во французском языке:			
	ton (couleur)	***тон (цвет)***	
	EM nuance de couleur	***см***	***оттенок цвета***
	ton (musique)	***тон (музыка)***	
	EM tonalité (musique):	***см***	***тональность***
В русском языке:			
	высота (звук)		
	см	***частота звука***	
	тон (звук)		
	см	***музыкальный звук***	
	тон (цвет)		
	см	***оттенок цвета.***	

9 Межъязыковая эквивалентность

Раздел 9 не включен в настоящий стандарт, поскольку вопросы разработки многоязычных тезаурусов рассмотрены в [11].

10 Отношения понятий

10.1 Введение

Кроме отношения эквивалентности имеется много других типов отношений, которые могут быть представлены в тезаурусе. Если отношение эквивалентности в одноязычном тезаурусе действует только между дескрипторами и аскрипторами, то отношения, устанавливаемые в настоящем разделе, действуют между понятиями (см. также модель данных в разделе 15). Такие отношения вводятся в тезаурус только при условии, что они парадигматические, т. е. имеют силу для всего разнообразия контекстов.

Примечание — Отношения между понятиями трудно отобразить, поскольку понятия существуют только в мыслях, и потому обычный метод состоит в том, чтобы показывать такие отношения как применяемые к дескрипторам, представляющим соответствующие понятия. Равным образом метки BT, NT, RT, которые отсылают к вышестоящим (более широким), нижестоящим (более узким) и ассоциативным терминам соответственно, могут рассматриваться как отсылающие к более широким, более узким и ассоциативным понятиям.

10.2 Иерархическое отношение

10.2.1 Общие положения

Иерархическое отношение устанавливается между двумя понятиями, если предмет одного понятия полностью включен в предмет другого понятия. Это отношение формируется в виде шкалы уровней иерархии, где вышестоящее понятие представляет собой род (класс) или целое, а подчиненное понятие является членом, видом или частью вышестоящего.

Соответственно отношения обозначают следующими метками, которые пишутся в тезаурусных статьях в виде префикса к вышестоящему или нижестоящему дескриптору.

В английском языке:

BT (broader term) — префикс к вышестоящему термину.

NT (narrower term) — префикс к нижестоящему термину.

В русском языке:

в (выше) — префикс к вышестоящему термину

н (ниже) — префикс к нижестоящему термину.[1)]

Пример —

animals		***животные***
NT	***mammals***	***н млекопитающие***
...		...
mammals		***млекопитающие***
BT	***animals***	***в животные.***

Иерархические отношения могут быть одного из трех типов:

a) родовидовое отношение;

b) отношение целое — часть (партитивное);

c) отношение множество — элемент (отношение членства).

Каждое из этих отношений выстраивает иерархии, которые подлежат проверке через обращение к соответствующим типам понятий, например, перечисленным в п. 5.1.2. Каждое подчиненное понятие должно принадлежать к той же категории, что и его вышестоящее понятие, т. е. и широкий, и узкий оба термина должны представлять либо вещь, либо действие, либо свойство и т. д. (см. дополнительно фасетный анализ в разделе 11).

Пример —

a) «металлы» (класс материалов) и «литье» (действие) представляют различные типы понятий и следовательно не могут быть связаны иерархически;

b) «металлы» и «золото» — оба представляют материалы и следовательно могут иметь иерархическую связь.

[1)] Для других языков [7].

Эти три типа иерархии можно явно различать в тезаурусе при использовании соглашений, описанных в 10.2.2, 10.2.3 и 10.2.4. Но это потребует дополнительной работы, и ее сложность должна быть оправдана ожидаемыми преимуществами.

Основная функция иерархических отношений состоит в помощи индексаторам и специалистам, ведущим поиск, в выборе надлежащего уровня специфичности. Поиск может быть расширен или сужен переходом по шкале иерархии вверх и вниз. Обычная техника увеличения полноты поиска называется «поисковым взрывом» или «взрывным поиском», при этом поиск расширяется за счет включения в поисковое предписание всех нижестоящих терминов первоначально выбранного дескриптора [см. 16.3.3, перечисления e) и f), 16.3.4, перечисление c)]. Тщательное следование указаниям пунктов с 10.2.2 — 10.2.5 обеспечит выдачу только тех материалов, которые соответствуют значению вышестоящего понятия (первоначально выбранного дескриптора).

Примечание — Взрывной поиск в отличие от расширенного поиска не распространяется на ассоциативные термины (см. 10.3)

10.2.2 Родовидовые отношения

10.2.2.1 Родовидовые отношения — это связь класса или категории со своими членами или видами. Дополнительно к проверке на состоятельность по 10.2.1 это отношение также подлежит логическому тесту «все и некоторые», как это показано на рисунке 2.

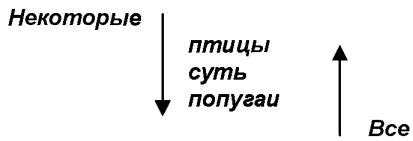


Рисунок 2 — Отношение, удовлетворяющее тесту «все и некоторые»

На рисунке 2 проиллюстрировано, что некоторые члены класса «птицы» известны как «попугаи», а все «попугаи» по определению независимо от контекста рассматриваются как «птицы». Этот тест обычно подтверждает, что такие термины как «попугаи» не подчинены классу «домашние животные», поскольку не все «попугаи» являются «домашними». На рисунке 3 показано отношение этой пары терминов.

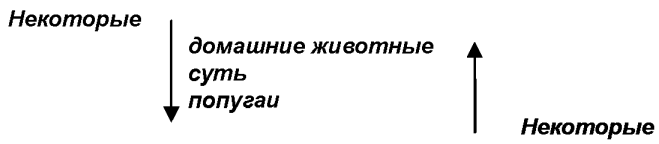


Рисунок 3 — Отношения, не удовлетворяющие тесту «все и некоторые»

На рисунке 3 показано, что только некоторые члены класса «домашние животные» являются «попугаями» и только некоторые «попугаи» являются «домашними животными».

Примечание — Работе на тему «попугаи как домашние животные» при индексировании должны быть присвоены два дескриптора, которые представляют оба эти класса.

10.2.2.2 Такое соотношение может не выполняться в специальном тезаурусе, посвященном «домашним животным», в котором попугаи рассматриваются только в данном аспекте. Тогда термин «попугаи» может быть подчинен классу «домашние животные» в одной и той же иерархии. Такие упрощения действительности следует принимать с осторожностью, особенно в контексте информационных сетей, в которых записи из одной системы могут соседствовать с записями из других систем. Для взаимодействия систем установленные отношения должны иметь универсальную значимость.

10.2.2.3 Ссылки **BT/NT** <u>в английском языке и **в/н** в русском</u>, как правило, достаточны для указания на родовидовые отношения, но при необходимости могут применяться более точные обозначения:

BTG (Broader term generic) = **вр** (вышестоящий термин родовой)

NTG (Narrower term generic) = **нв** (нижестоящий термин видовой).

Пример —	*rats*		*крысы*	
	BTG	*rodents*	*вр*	*грызуны*
	...		*...*	
	rodents		*грызуны*	
	NTG	*rats*	*нв*	*крысы.*

10.2.3 Иерархическое отношение «целое — часть»

Иерархическое отношение целое — часть охватывает ограниченный круг ситуаций, в которых часть объекта или системы принадлежит исключительно одному определенному целому. Это относится к следующим четырем главным классам терминов.

a) Системы и органы тела:

Пример — кровеносная система
кровеносные сосуды
артерии
вены.

b) Географические области:

Пример — Канада
Онтарио
Оттава
Торонто.

c) Дисциплины или тематические области:

Пример — наука
биология
ботаника
зоология.

d) Иерархические социальные структуры:

Пример — армия
корпус
дивизия
батальон
полк.

Большинство других случаев отношений «целое — часть» не подлежат иерархическому связыванию, поскольку часть может принадлежать более, чем одному целому. Например, отношение BT/NT не следует устанавливать между терминами «велосипеды» и «колеса», поскольку колесо может быть частью автомобиля, тачки и многих других изделий. Общий поиск по термину «велосипеды» привел бы к выдаче множества нежелательного материала, если бы поиск был расширен по всем типам колес. Однако в некоторых случаях части изделия принадлежат исключительно данному изделию, по крайней мере, в сфере применения тезауруса. Может оказаться целесообразным установление связей BT/NT, например, между терминами «fireplaces» и «hearths», «лук» и «тетива». Это обычно не приводит к противоречию и содействует общему поиску по вышестоящему термину. Но такая практика не рекомендуется для случая компонентов сложных машин, для которых более разумным решением будет образование вышестоящего термина «компоненты устройства», под которым будут перечислены различные компоненты в качестве видовых терминов (см. 7.5.2).

10.2.3.1 Пометы BT/NT обычно достаточны для указания на партитивное отношение, но могут также применяться более точные пометы:

BTP (broader term partitive) = **вц** (вышестоящий термин целого)

NTP (narrower term partitive) = **нч** (нижестоящий термин части)

Пример —
центральная нервная система
вц нервная система
... ...
нервная система
нч центральная нервная система.

10.2.4 Отношение множество — элемент

10.2.4.1 Отношение множество — элемент связывает общее понятие, такое как класс вещей или событий, с элементами этого класса, которые часто бывают представлены именами собственными (см. 6.6.9).

Пример —
горные массивы — класс
Альпы
Гималаи — элементы класса.

В этом примере названия «Альпы» и «Гималаи», присвоены подчиненным позициям в иерархии. Однако они не являются ни видами, ни частями понятия «горные массивы», а представляют отдельные элементы.

10.2.4.2 Пометы BT/NT обычно достаточны для указания на отношение множество — элемент, но кроме этого могут применяться также и точные пометы:

BTI (broader term instantial) = **вм** (вышестоящий термин множества);

NTI (narrower term instantial) = **нэ** (нижестоящий термин элемента).

Пример —

Париж
вм столицы
... ...
столицы
нэ Париж.

10.2.5 Полииерархические отношения

Некоторые понятия логически принадлежат одновременно более чем одной группе или классу. В таких случаях должна быть установлена прямая связь со всеми соответствующими широкими понятиями, а структура тезауруса тогда называется полииерархической. Она противопоставляется моноиерархической структуре, которая допускает установление для каждого понятия только одной действительной связи BT.

Пример —



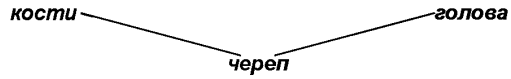
В примере термин «органы» присвоен подчиненной позиции родовидового отношения к двум вышестоящим терминам. В других случаях полииерархические связи могут быть основаны на отношении целое — часть.

Пример —



В некоторых случаях полииерархические отношения могут быть основаны на различных типах отношений.

Пример —



В примере связь между терминами «кости» и «череп» основана на родовидовом отношении (череп это вид костей), а связь между терминами «голова» и «череп» основана на иерархическом отношении целое — часть (череп есть часть головы).

Когда одно понятие имеет более одного вышестоящего понятия, должны быть указаны связи со всеми соответствующими терминами. Факультативно может быть показано различие типов отношений посредством соответствующих помет.

Пример —

череп		***или***	***череп***	
в	***кости***		***вр***	***кости***
	голова		***вц***	***голова***

10.2.6 Несимметричные структуры во многоязычных и мультикультурных тезаурусах

Обычно все языковые версии многоязычного тезауруса имеют одинаковую структуру иерархии понятий. Однако могут возникать сложности установления иерархии понятий, когда две или более социальные общности, различающиеся по культуре, пользуются одним тезаурусом, особенно если понятия и термины, известные одной общности, не имеют однозначных соответствий у другой. Чтобы согласо-

вать культурные и языковые различия и обеспечить равный статус языковым версиям одного тезауруса могут быть введены несимметричные структуры. Подходы к построению многоязычных тезаурусов см. в 13.3.3.

Примечание — Если применяются несимметричные структуры, то модель данных, описанная в п. 15, должна быть модифицирована.

10.3 Ассоциативное отношение

10.3.1 Общие положения

Ассоциативные отношения охватывают ассоциации между парами понятий, которые не связаны иерархически, но связаны по смыслу или по единству природы в такой мере, что связь между ними должна быть указана в тезаурусе, поскольку благодаря наличию этой связи при проведении индексирования или поиска удается предложить альтернативный термин. Это отношение обозначается пометой «**RT**» (related term = родственный термин) в английском языке и пометой «**а**» в русском языке, которые указываются взаимно для каждого из связанных терминов.

Пример —

птицы
а орнитология
... ...
орнитология
а птицы.

Общее значение ассоциативных отношений таково: если используется одно из понятий, в рамках обычных для пользователей тезауруса представлений другое понятие подразумевается. Более того, один из терминов часто необходим при определении или объяснении другого. Например, термин «птицы» составляет необходимую часть объяснения термина «орнитология».

В частности, важно устанавливать ассоциативное отношение между понятиями с пересекающимся охватом.

10.3.2 Термины и понятия с пересекающимися значениями

10.3.2.1 В повседневном употреблении некоторые термины в одних контекстах могут заменять друг друга, а в других — нет. Например, термины «ships — суда» и «boats — лодки» могут быть употреблены для обозначения различных понятий, поскольку, хотя они имеют много общего, ни один из них не покрывает объем другого понятия. Если два таких понятия введены в тезаурус и представлены различными дескрипторами, то целесообразно установить между ними ассоциативную связь. При поиске это напомнит о том, что возможно следует использовать оба дескриптора, чтобы быть уверенным в отыскании всего релевантного материала.

Пример —

boats		***лодки***	
RT	***ships***	***а***	***суда***
... ...		***... ... ...***	
ships		***суда***	
RT	***boats***	***а***	***лодки.***

10.3.2.2 Дескрипторы с пересекающимися значениями могут быть близкородственными терминами, «братьями», т. е. они имеют общий вышестоящий термин. Но нет необходимости связывать ассоциацией все близкородственные термины. Например, не нужно связывать ассоциацией термины «лошади» и «ослы», потому что они имеют общий вышестоящий термин «лошадиные», поскольку значения терминов «лошади» и «ослы» не пересекаются.

10.3.3 Другие случаи ассоциативных связей

Достаточные основания для ассоциативной связи имеются тогда, когда один термин устойчиво вызывает представление о другом термине. Нижеследующие примеры представлены только как типичные ситуации отношений, встречающиеся на практике.

а) Дисциплина или область исследования и изучаемые объекты или явления:

Примеры —

1 лесное хозяйство
а лес
... ...

лес
a лесное хозяйство
2 нейрология
a нервная система
... ...
нервная система
a нейрология.

b) Операция или процесс и его агент или инструмент:

Примеры —
1 регулирование температуры
a термостаты
... ...
термостаты
a регулирование температуры
2 уголовный розыск
a следователи
... ...
следователи
2 a уголовный розыск.

c) Действие и результат действия:

Примеры —
1 weaving *ткачество*
RT cloth *a ткань*
... ... *... ... ...*
cloth *ткань*
RT weaving *a ткачество.*

2 пахота
a плуги
... ...
плуги
a пахота.

d) Действие и то, на что оно направлено:

Примеры —
1 сбор урожая
a урожай
... ...
урожай
a сбор урожая
2 арест
a арестанты
... ...
арестанты
арест.

e) Объекты или материалы и их определяющие свойства:

Примеры —
1 магниты
a ферромагнетизм
... ...
ферромагнетизм
a магниты

2 яды
a токсичность
... ...
токсичность
a яды.

f) Изделия и их части, если они не пригодны для отношения целое — часть (см. 10.2.3):

Пример —

оптические инструменты
a линзы
… … …
линзы
a оптические инструменты.

g) Понятия, связанные причинной связью:

Примеры —
1 bereavement
RT death
… … …
death
RT bereavement

2 болезни
a патогены
… … …
патогены
a болезни.

h) Объект или процесс и противодействующий агент:

Примеры —
1 растения
a гербициды
… … …
гербициды
a растения
2 возгорание
a огнетушители
… … …
огнетушители
a возгорание.

i) Понятие и единица его измерения:

Пример —

электрический ток
a амперметры
… … …
амперметры
a электрический ток.

j) Сложный термин и опорное существительное словосочетания, если соответствующие два понятия не находятся в точном иерархическом отношении:

Примеры —
1 модели кораблей
a корабли
… … …
корабли
a модели кораблей.
2 ископаемые пресмыкающиеся
пресмыкающиеся
… … …
пресмыкающиеся
a ископаемые пресмыкающиеся.

k) Организм или субстанция, произошедшая или произведенная из другой:

Примеры —
1 мулы
а ослы
... ...
ослы
а мулы.
2 латунь
а медь
... ...
медь
а латунь.

10.3.4 Многоязычные и мультикультурные тезаурусы

Для того чтобы обеспечить достаточность ассоциативных связей во многоязычном тезаурусе, должны быть учтены потребности в установлении связей пользователей всех языковых версий.

Как уже описано в 9.1, 9.2 и 10.2.6, сообщества с различными культурами или различными языками, пользующиеся одним и тем же тезаурусом, иногда нуждаются в таких понятиях и связях понятий, которые незнакомы другим. Когда для удовлетворения этих потребностей вводятся несимметричные иерархические структуры, как правило, оказываются необходимыми также и несимметричные ассоциативные связи.

Примечание — Если допускаются несимметричные структуры, модель, описанная в разделе 15, должна быть модифицирована.

10.4 Пользовательские отношения

Отношения эквивалентности, иерархии и ассоциации, описанные в 8, 10.2 и 10.3, хорошо определены, широко используются и обычно рассматриваются как вполне адекватные в контексте использования человеком для получения сведений о вышестоящих, нижестоящих и связанных терминах при выборе способа выражения своей информационной потребности. Однако иногда возникает необходимость в дополнительных или более точных типах отношений.

Таким же образом, как иерархические отношения разрешается подразделять на отношения род — вид, целое — часть и множество — элемент, так и отношения эквивалентности и ассоциации разрешается подразделять, если это необходимо для специального использования. Например, для аббревиатур и акронимов может быть введено отдельное отношение эквивалентности. Для их обозначения должны быть назначены пометы, отличные от USE/UF, например FT/AB от «Full Term» и «Abbreviation», или для русского языка — **пф/кф** от «полная форма» и «краткая форма». Аналогично для ассоциативных отношений можно ввести подразделение на такие два типа отношений как причина и следствие, обозначаемые такими пометами как CAUSE/EFFECT, или ПРИЧИНА/СЛЕДСТВИЕ.

Прежде чем предпринять попытку установления более специфичных отношений, разработчик тезауруса должен убедиться, что уточненные отношения будут действительно необходимы, а не останутся просто привлекательным интеллектуальным конструктом. Предназначение тезауруса состоит в том, чтобы служить определенному коллективу пользователей в определенных обстоятельствах. И выгоду, получаемую от введения дополнительных отношений, может перевесить дополнительная сложность восприятия этих отношений пользователями. Однако проведение дополнительной работы можно считать оправданным, если тезаурус предполагается использовать в функции онтологии или во взаимодействии с ней. (В онтологиях обычно прорабатываются более специфичные и четко определенные отношения).

Примечание — См. рекомендации по этим вопросам в настоящем стандарте.

В случае, когда в тезаурусе имеются специфические особенности, требуемые заказчику, важно чтобы пользователь был осведомлен об этом и, чтобы использование такого специализированного тезауруса совместно с обычным тезаурусом не вызывало недоумения. Введение нестандартных помет или символов влечет за собой опасность потери совместимости при взаимодействии данного тезауруса с другими. Этот риск может быть уменьшен, если предоставляется гарантия того, что новые отношения являются истинными подвидами стандартных типов отношений.

11 Фасетный анализ

Фасетный анализ полезен при установлении иерархических отношений, удовлетворяющих правилам 10.2, поскольку эти отношения действуют только между понятиями, принадлежащими одной и той же общей категории.

Выбор фасетов может варьировать в зависимости от тематики тезауруса, но на самом верхнем уровне обычно используют фундаментальные категории — ***объект, материал, действующий агент, действие, место, время*** и т. п. Там, где это целесообразно, эти фасеты могут быть разбиты на подфасеты до требуемого уровня дробности. Например, действия могут быть подразделены на нетранзитивные ***процессы*** (например, «созревание» или «ухудшение») и транзитивные ***операции*** (например, «отрезание» или «исправление»).

Примечание — Действие называется нетранзитивным, если действующий агент не воздействует на другой объект, в то время как агент транзитивного действия воздействует на объект действия. Так, созревание яблока — внутренний процесс, но при разрезании яблока нож воздействует на него как на объект операции.

Пример применения фасетного анализа при классификационном упорядочении приведен на рисунке 4, где показаны некоторые термины из общего класса «industries». Непосредственно под этим термином располагаются два подчиненных термина «agricultural industries — сельскохозяйственное производство» и «engineering industries — промышленное производство». Один из фасетов, показанных под «agricultural industries» является «products — продукты». В этом фасете показано только два уровня иерархии, за исключением термина «milk — молоко», который расписан так, что показано, как его нижестоящие термины сгруппированы в ряды, каждому из которых предшествует метка узла, указывающая на основание деления. Эти ряды представляют различные способы подразделения понятия «milk» — по жирности, по биологическому виду животных и т. п., на что указывает слово «by = по» в составе каждой метки узла. В отличие от этого метки узлов, которые вводят новые фасеты, такие как «people» и «products», не содержат слова «by».

Необходимо отметить различие в структуре иерархических отношений терминов, разделенных метками узлов. Если метка показывает основание деления понятия, все термины, следующие за нею, обозначают подчиненные видовые понятия для вышестоящего термина. В то время как там, где метка узла вводит новый фасет, следующие термины обычно не соответствуют видовым понятиям предыдущего термина. Например, «farm managers = менеджеры ферм» и «cereal products = зерновые продукты» и т. п. не являются видами «agricultural industries — сельскохозяйственное производство».

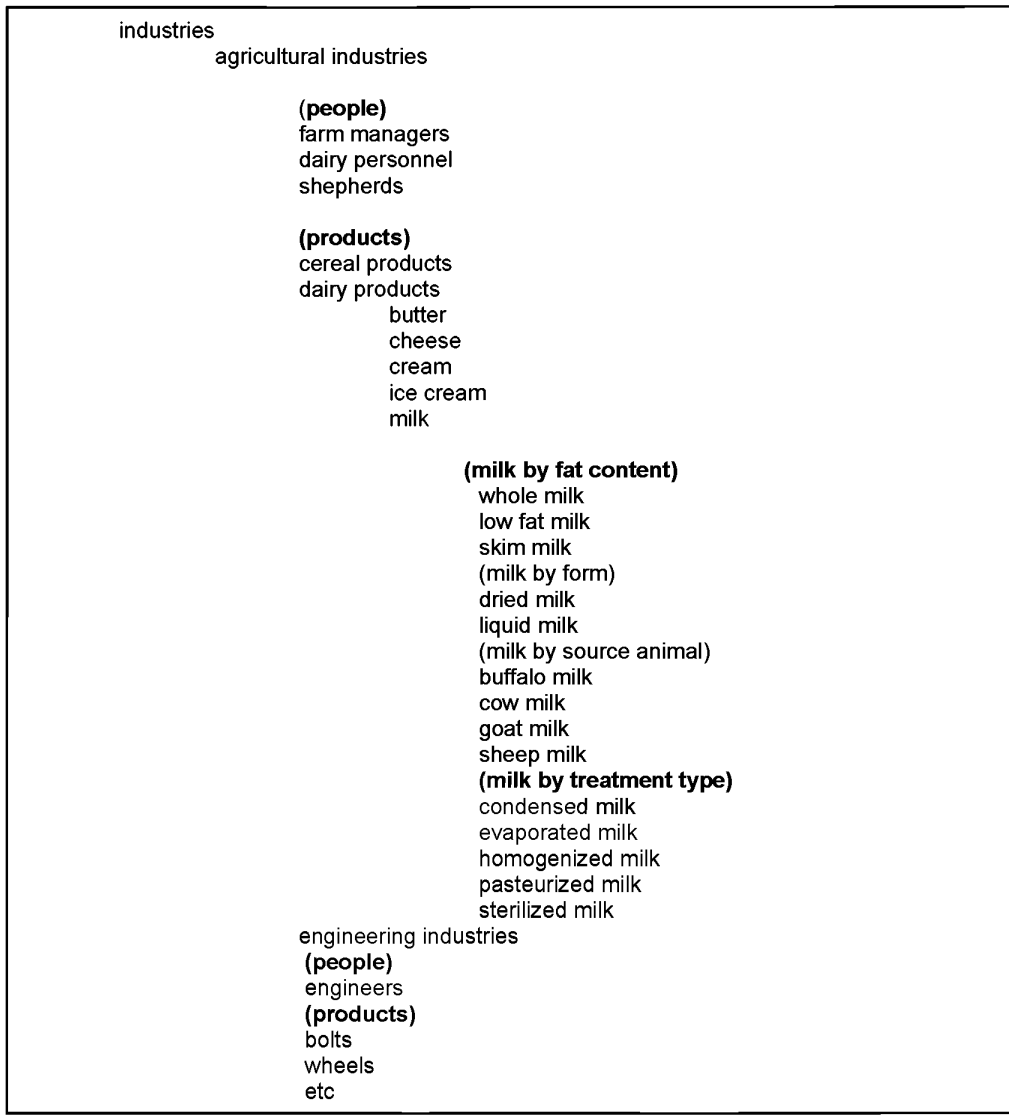
Метки узлов не являются терминами тезауруса. Они служат только для систематизации выдачи и они не могут быть членами отношений, описанных в разделе 8 — 10. Чтобы избежать недоразумений, метки узлов должны быть выделены шрифтом, отличным от шрифта терминов. Обычно для этого используют курсив и скобки, например как на рисунке 4. Некоторые другие пути применения фасетного анализа проиллюстрированы в 12.2.4 и 12.2.5.

Термины, представляющие понятия одного ряда, могут быть представлены как в алфавитном, так и в систематическом порядке. Алфавитная последовательность целесообразна тогда, когда нет других очевидных путей группировки понятий. Систематический порядок целесообразен тогда, когда он использован большей частью пользователей или когда он помогает прояснить значение терминов. В примере с электромагнитным излучением (приведенным ниже) виды излучения расположены в порядке возрастания длины волны, что может помочь индексаторам в выборе подходящего термина.

Пример —
Электромагнитное излучение
<по длине волны>
ультрафиолетовое излучение
видимый свет
инфракрасное излучение
микроволновое излучение
радиоволны.

Примечание — Этот пример использует слегка измененные, но приемлемые соглашения о метке узла: угловые скобки вместо круглых, и родительский термин не выписан в начале метки узла. Присутствие слова «по», однако, показывает, что метка узла определяет основание деления и расположения типов излучений.

В некоторых тезаурусах имена фасетов включены в качестве дескрипторов и рассматриваются как термины высшего уровня, под которыми может быть показана полная иерархия терминов (см. 12.2.4). В других случаях может не существовать полного указателя всех фасетов, а фасетные имена появляются только в качестве меток узлов, как это представлено на рисунке 4, или в качестве имен понятийных групп (см. 15.2.18 — 15.2.19).



EXAMPLE 1 «Produts» is an example of a node label that indicates changes of facet.
EXAMPLE 2 «Milk by source animal» is an example of a node label indicates characteristics of division of arrays.

Рисунок 4 — Частично развернутый классификационный указатель класса «industries» с метками узлов, показывающими смену фасетов и оснований деления рядов терминов

12 Представление и оформление тезауруса

12.1 Общие положения

12.1.1 Использование базы данных для ведения указателей информации

В компьютерной системе тезаурус обычно представлен в структуре базы данных (см. раздел 15), где каждый отдельный термин и каждая связь пары терминов записаны только один раз. Это обеспечивает постоянство отношений и позволяет выделять и показывать термины различными способами.

12.1.2 Обзор стилей представления тезауруса

Представление тезауруса на человекочитаемом носителе, на экране или в печатной форме, возможно с помощью следующих приемов:

а) единая запись — это основная форма представления тезауруса, включающая статьи дескрипторов и аскрипторов и выборочно некоторые или все их отношения, коды и примечания;

b) алфавитное упорядочение, обеспечивающее доступ к понятиям по словам, которыми пользователи первоначально выразили понятия. В печатном тезаурусе такая форма служит одним из указателей, а при компьютерном представлении она способна дополнять функцию прямого поиска;

c) иерархическое упорядочение на основе отношений выше/ниже позволяет расширять и уточнять понятия при индексировании и поиске;

d) классификационное упорядочение позволяет просматривать предметную область по связям понятий;

e) графическое отображение показывает термины и их отношения наглядным образом;

f) пермутационный указатель облегчает поиск слов, входящих в состав многословных терминов.

Эти различные приемы дополняют друг друга. Важно, чтобы был обеспечен алфавитный доступ, либо путем прямого поиска, либо по алфавитному указателю, как описано в перечислении b). Должен иметься по крайней мере один из указателей — c) или d), дающий систематический обзор тезауруса. Дополнительные данные о каждом понятии, такие как лексические примечания и отношения с другими понятиями, могут быть показаны в каждом из этих указателей. Тезаурусы могут сильно отличаться в оценке относительной важности и функций различных указателей, как в части упорядочения, так и в части представляемой в этих указателях информации. Графический указатель может предоставлять полезный общий обзор области знания, но этот указатель сложен для создания и ведения, и его структура затемняется, если он содержит слишком много дополнительных данных. Пермутационный указатель служит как справочник всех слов во всех терминах (дескрипторах и аскрипторах), для этого можно использовать форматы KWIC (key word in context — ключевое слово в контексте) или KWOC (key word out of context — ключевое слово вне контекста). Этот указатель полезен для печатных тезаурусов, но обычно не требуется для автоматизированных систем, в которых возможен поиск по последовательности букв.

12.1.3 Нотация и связи между указателями

Алфавитный указатель в печатной форме должен показывать место термина в других указателях с помощью либо номера строки, либо системы обозначений, либо с помощью указания термина высшего уровня в иерархии данного термина (если иерархии тезауруса не слишком обширны). Для связи с графическим указателем от алфавитного указателя может понадобиться система координат. При выводе указателей на экран компьютера могут не понадобиться отображаемые символы связей, поскольку автоматически сформированные гиперссылки должны обеспечивать легкое переключение от одного указателя к другому, сохраняя при этом позиционирование на рассматриваемом термине.

В классификационном указателе нотация может отображать классификационную структуру и быть рассчитана на создание кодов предкоординированных классов, отражающих сложные понятия. Любая система нотации должна допускать введение нового понятия в любой точке классификационного упорядочения терминов.

12.2 Альтернативы стилей отображения тезауруса

12.2.1 Общие положения

Существует слишком много возможностей разных стилей представления тезауруса, что не позволяет их все здесь проиллюстрировать. В пунктах от 12.2.2 — 12.2.6 приведены примеры основных типов указателей. Чтобы облегчить сравнение и сопоставление, они иллюстрированы общим набором терминов, выражающих понятие «cameras = фото- и кинокамеры» и связанные с ним. Иллюстрации подобраны так, чтобы содержать все основные типы отношений, а также некоторые факультативные элементы, такие как метки узлов и нотация, которые представляют сложность для ведения. Там, где присутствует нотация, использована простая система номеров строк.

Примечание — В этом разделе не рассматриваются многоязычные тезаурусы. Они описаны в 12.3.

12.2.2 Представление одной словарной статьи

Представление одиночной записи о понятии или дескрипторе должно обычно начинаться с дескриптора, за которым следуют другие поля словарной статьи в следующем порядке:

a) **CC** — код понятия или обозначение, указывающее место термина/понятия в иерархии;

b) **SN** — лексическое примечание;

c) **UF** — ссылки на эквивалентные аскрипторы;

d) **TT** — ссылки на термины высшего уровня;

e) **BT** — ссылки на вышестоящие термины;

f) **NT** — ссылки на нижестоящие термины;

g) **RT** — ссылки на ассоциированные термины;

h) **DEF** — определение дескриптора (см. 6.2.3);

i) **HN** — историческое примечание (см. 6.2.4);

j) **SC** — предметная категория или другая понятийная группа, к которой относится понятие.

Здесь ссылки **SN** и **UF** даны в начале списка, поскольку они проясняют значение понятия. За ними следуют отношения к другим понятиям. **DEF** и **HN** идут в конце, поскольку рассматриваются как административные поля, используемые больше редакторами, чем при поиске или индексировании. Их возможно также указывать сразу после лексического примечания; тем самым все текстовые поля будут собраны вместе. Однако важно обеспечить, чтобы **DEF** и **SN** не смешивались в том случае, если и то и другое поле присутствует.

Для аскрипторных статей порядок полей следующий:

- **USE** — ссылка на соответствующий дескриптор (предпочтительный термин);
- **DEF** — определение аскриптора (непредпочтительного термина) (см. 6.2.3);
- **HN** — историческое примечание (см. 6.2.4).

Для отдельных приложений выбор и последовательность элементов могут быть изменены. Например, редактор тезауруса может захотеть для собственного использования включить некоторые административные поля (см., например, 15.2.4, 15.2.9, 15.2.14).

12.2.3 Алфавитный указатель

12.2.3.1 В алфавитном указателе все термины (дескрипторы и аскрипторы) приводятся в единой алфавитной последовательности. Дескрипторы и аскрипторы должны по возможности различаться типографским способом, так чтобы непредпочтительный статус аскрипторов был очевиден. (В печатном издании для этого можно применять курсив). В некоторых случаях будет достаточен простой список терминов. В большинстве случаев в указатель включают полные записи каждого термина согласно 12.2.2.

12.2.3.2 Пример такого указателя приведен на рисунке 5. Кроме указателя, обеспечивающего выдачу единой тезаурусной статьи, этот указатель является простейшим типом представления тезауруса с точки зрения его формирования и ведения. Чаще всего принимается соглашение указывать только непосредственно вышестоящие и нижестоящие термины; при этом экономится место, но другие уровни иерархии перестают быть очевидными.

12.2.3.3 Можно показывать и больше одного уровня иерархии, но тогда следует указывать номер уровня вышестоящих и нижестоящих терминов, как делается в примере, приведенном ниже. Однако если использовать это соглашение, то место, необходимое для показа всех уровней иерархии всех терминов может оказаться слишком большим.

Пример —

полорогие

BT1 жвачные

BT2 млекопитающие

BT3 позвоночные

NT1 буйволы

NT1 крупный рогатый скот

NT2 мясной скот

NT2 молочный скот

NT1 яки.

35 mm cameras
CC: H012
BT: film cameras

aqualungs
CC: D002
BT: diving equipment

camera accessories
CC: H002
BT: photographic equipment
NT: flash guns
light meters
tripods
RT: cameras

camera components
CC: H006
BT: cameras and camera components
NT: camera lenses
camera viewfinders

camera lenses
CC: H007
BT: camera components

camera viewfinders
CC: H008
BT: camera components

cameras
CC: H009
BT: cameras and camera components
NT: digital cameras
film cameras
instant picture cameras
plate cameras
reflex cameras
special-purpose cameras
RT: camera accessories
photography

cameras and camera components
CC: H005
BT: photographic equipment
NT: camera components
cameras

colour
CC: A204
BT: optical properties

contrast
CC: A205
BT: optical properties

digital cameras
CC: H010
UF+: *underwater digital cameras*
BT: cameras

divers
CC: D008
BT: people
RT: diving

diving
CC: D000
BT: fields of work
RT: divers
diving equipment

diving equipment
CC: D001
BT: equipment
NT: aqualungs
diving suits
face masks
swimming fins
underwater cameras
RT: diving

diving suits
CC: D003
BT: diving equipment
NT: dry suits
wet suits

dry suits
CC: D004
BT: diving suits

equipment
CC: A001
NT: diving equipment
electrically-powered equipment
fixed equipment
human-powered equipment
photographic equipment
physics equipment
portable equipment

exposure meters
USE: light meters

face masks
CC: D006
BT: diving equipment

fields of work
CC: A300
NT: diving
photography
physics

film cameras
CC: H011
BT: cameras
NT: 35 mm cameras
medium format cameras
miniature cameras

instant picture cameras
CC: H015
SN: Cameras which produce a finished print directly
UF: *Polaroid cameras®*
BT: cameras

people
CC: A100
NT: adults
children
divers
infants
models (people)
photographers
physicists

photographic equipment
CC: H001
BT: equipment
NT: camera accessories
cameras and camera components
RT: photography

photography
CC: H000
BT: fields of work
RT: cameras
photographers
photographic equipment

physical properties
CC: A202
BT: properties
NT: pressure
temperature

physicists
CC: P005
BT: people
RT: physics

physics
CC: P000
BT: fields of work
RT: physicists

Polaroid® cameras
USE: instant picture cameras

pressure
CC: A208
BT: physical properties

properties
CC: A200
NT: physical properties

reflex cameras
CC: H017
SN: Cameras in which the image is reflected on to a glass screen for composing and focusing.
BT: cameras
NT: single lens reflex cameras
twin lens reflex cameras

single lens reflex cameras
CC: H018
UF: *SLR cameras*
BT: reflex cameras

SLR cameras
USE: single lens reflex cameras

special-purpose cameras
CC: H021
BT: cameras
NT: stereo cameras
underwater cameras

stereo cameras
CC: H022
BT: special-purpose cameras

swimming fins
CC: D007
BT: diving equipment

temperature
CC: A209
BT: physical properties

tripods
CC: H004
BT: camera accessories

twin lens reflex cameras
CC: H019
BT: reflex cameras

underwater cameras
CC: D028, H023
UF+: *underwater digital cameras*
BT: special purpose cameras
diving equipment
RT: diving

underwater digital cameras
USE: digital cameras
\+ underwater cameras

Рисунок 5 — «Алфавитный указатель тезаурусных терминов (некоторые термины пропущены для экономии места»)

12.2.4 Иерархический указатель

Иерархический указатель для одного языка формируют путем размещения каждого ряда нижестоящих терминов под их родительским вышестоящим термином. Другие отношения (такие как отношения эквивалентности и ассоциации) обычно в этом указателе не показывают, поскольку это сделало бы структуру указателя трудной для восприятия. В результате получается множество иерархических деревьев, возглавляемых различными терминами высшего уровня (не имеющими над собой вышестоящих).

В противном случае после произведения фасетного анализа на верхнем уровне тезауруса, все термины одного фасета могут сгруппироваться в одну иерархию, возглавляемую меткой узла, содержащей имя фасета, как показано на рисунке 6. (Имена фасетов на рисунке 6 — «objects = объекты», «feld of work = область деятельности», «people = люди», «properties = свойства»).

Когда фасетный анализ применяется в такой форме, термины группируются в фасеты согласно базисным категориям тех понятий, которые эти термины представляют (см. 10.2.1 и раздел 11), несмотря на то, с какой областью или областями эти понятия обычно соотносятся.

При сравнении этого подхода с изложенным в 12.2.5.2, где первоначальная организация терминов идет по темам, и фасеты формируются внутри тем, можно отметить следующие преимущества первоначального проведения фасетного анализа:

a) менее вероятна необходимость серьезной перестройки указателя в случае, если некоторое понятие изменит свою тематическую область;

b) при взаимодействии с другими информационными центрами можно ожидать достижения большего уровня согласованности.

Однако необходимо отметить и следующие недостатки:

a) организация по фасетам разбрасывает понятия, связанные внутри своей тематики, по разным группам;

b) понятийный базис, на котором построен тезаурус, делается менее очевидным;

c) если тезаурус охватывает все или широкий диапазон дисциплин, то вряд ли объединение элементов и процессов различных дисциплин приведет к образованию разумной последовательности терминов. Следовательно, такой вид указателя целесообразно использовать только в случае тезаурусов, посвященных одной определенной дисциплине.

(objects)
equipment
. *<equipment by portability>*
. . fixed equipment
. . portable equipment
. *<equipment by power source>*
. . electrically-powered equipment
. . human-powered equipment
. *<equipment by application>*
. . diving equipment
. . . aqualungs
. . . diving suits
. . . dry suits
. . . wet suits
. . . face masks
. . . swimming fins
. . . underwater cameras
. . photographic equipment
. . . camera accessories
. . . . flash guns
. . . . light meters
. . . . tripods
. . . cameras and camera components
. . . . camera components
. . . . . camera lenses
. . . . . camera view finders
. . . . cameras
. . . . . *<cameras by imaging technique>*
. . . . . digital cameras
. . . . . film cameras
. . . . . *<film cameras by film size>*
. . . . . 35 mm cameras
. . . . . medium format cameras
. . . . . miniature cameras
. . . . . instant picture cameras
. . . . . plate cameras
. . . . . *<cameras by viewing method>*
. . . . . reflex cameras
. . . . . single lens reflex cameras
. . . . . twin lens reflex cameras
. . . . . viewfinder cameras
. . . . . special-purpose cameras
. . . . . stereo cameras
. . . . . underwater cameras

(objects) (cont.)
equipment *(cont.)*
. *<equipment by application> (cont.)*
. . physics equipment
. . . optical instruments
. . . . light meters
. . . . microscopes

(fields of work)
fields of work
. diving
. photography
. physics

(people)
people
. *<people by age>*
. . infants
. . children
. . adults
. *<people by profession>*
. . divers
. . models (people)
. . photographers
. . physicists

(properties)
properties
. physical properties
. . optical properties
. . . colour
. . . contrast
. . . luminance
. . . luminosity
. . pressure
. . temperature

Рисунок 6 — Иерархический указатель тезаурусных терминов, упорядоченный по фасетам

12.2.5 Классификационный указатель

12.2.5.1 Упорядочивание терминов тезауруса по тематическим полям

Организация терминов тезауруса по дисциплинам или тематическим областям подобна подходу, используемому при разработке библиотечных классификаций, когда все начинается с распределения универсума понятий по классам основных дисциплин. Применительно к тезаурусам категории понятий сначала необходимо сгруппировать так, чтобы это соответствовало интересам пользователей. Эта техника применяется к тезаурусам, охватывающим некоторый диапазон тематических областей. Те понятия, которые принято связывать с некоторой областью, например «искусство», собираются вместе и эксплицитно отграничиваются от понятий, принадлежащих другим областям интересов, например таких как «экономика» или «физика».

Этот подход также может быть применен и к таким тезаурусам, которые охватывают только одно тематическое поле. Так, лексику тезауруса, посвященного медицине, сначала распределяют по таким подполям, как «хирургия», «физиология», «анатомия» и т. п. Кроме того, в качестве расширения ядерных понятий в тезаурус могут быть включены и понятия из таких других дисциплин как «управление», «право», «обработка данных», которые следует отделять от медицинских терминов.

Предметный подход имеет то преимущество, что позволяет организовать понятия в группы, которые в целом соответствуют образу мыслей пользователей. Трудность заключается в том, что одно и то же понятие часто используется в нескольких различных областях, например, если тезаурус охватывает как лесное хозяйство, так и агрономию, то он будет содержать много понятий, таких как «растение», которые встречаются и там, и там. Такие понятия следует либо повторять в каждом поле, где они используются, либо включать в отдельный список общих понятий, не ограниченных использованием в каком-либо одном поле.

Группирование понятий по тематическим полям с алфавитной или тематической упорядоченностью родственных терминов и с расположением узких понятий под соответствующими широкими, соблюдая абзацный отступ, приводит к построению такого указателя, который хотя и выглядит как иерархический, но в котором, подчиненные термины могут не подчиняться правилам для иерархических отношений, описанным в п.10.2. Следовательно, метки узлов, дающие имена фасетам, должны быть введены в такой указатель для того, чтобы отметить, где следуют различные фасеты и где отношения не соответствуют иерархии по правилам раздела 11.

12.2.5.2 Фасетная организация терминов тезауруса внутри предметных областей

На практике различные подходы, описанные в 12.2.4 и 12.2.5.1 часто комбинируют, как, например, в случае, когда тезаурус сначала организуют по тематическим полям, а далее подразделяют по фасетам.

На рисунке 7 проиллюстрирован комбинированный подход на примере тезауруса, который включает тематику подводного плавания, фотографии и физики. В каждом из этих полей показаны только те дескрипторы, которые относятся только к этому полю. Термины, общие для всех трех полей, представлены в разделе общих терминов. После первичного деления по тематике следует деление по фасетам. На рисунке 7 каждый тематический раздел поделен на фасеты «объекты», «люди» и «свойства».

В печатном указателе часто бывает необходимо ограничить общий размер продукта путем выбора одного основного места для каждого термина. Термин может быть повторен и в других местах, но пространство экономится за счет исключения повторного перечисления нижестоящих терминов. Предполагается, что пользователь может посмотреть основную запись термина, где найдет полную иерархию. Так на рисунке 7 основное место термина «properties» есть A200, где показаны нижестоящие термины «chemical properties», «physical properties» и все другие. В разделе подводного плавания единственные физические термины, которые были включены, это «pressure» и «temperature». В фотографическом разделе полностью приведены только оптические свойства.

A000 general terms
(objects)
A001 . equipment
. . *<equipment by portability>*
A002 . . . fixed equipment
A003 . . . portable equipment
. . *<equipment by power source>*
A004 . . . electrically-powered equipment
A005 . . . human-powered equipment
. . *<equipment by application>*
. . . diving equipment (D001)
. . . photographic equipment (H001)
. . . physics equipment (P001)
(people)
A100 . people
. . *<people by age>*
A101 . . . infants
A102 . . . children
A103 . . . adults
. . *<people by profession>*
. . . divers (D008)
. . . models (people) (H025)
. . . photographers (H026)
. . . physicists (P005)
(properties)
A200 . properties
A201 . . chemical properties
A202 . . physical properties
A203 . . . optical properties
A204 . . . . colour
A205 . . . . contrast
A206 . . . . luminance
A207 . . . . luminosity
A208 . . . pressure
A209 . . . temperature

(fields of work)
A300 . fields of work
. . diving (D000)
. . photography (H000)
. . physics (P000)

D000 diving
(objects)
. equipment (A001)
D001 . . diving equipment
D002 . . . aqualungs
D003 . . . diving suits
D004 . . . . dry suits
D005 . . . . wet suits
D006 . . . face masks
D007 . . . swimming fins
. . . underwater cameras (H023)
(people)
. people (A100)
D008 . . divers
(properties)
. properties (A200)
. . physical properties (A202)
. . . pressure (A208)
. . . temperature (A209)

H000 photography
(objects)
. equipment (A001)
H001 . . photographic equipment
H002 . . . camera accessories
H003 . . . . flash guns
. . . . light meters (P003)
H004 . . . . tripods
H005 . . . cameras and camera components
H006 . . . . camera components
H007 . . . . . camera lenses
H008 . . . . . camera viewfinders
H009 . . . . cameras
. . . . . *<cameras by imaging technique>*
H010 . . . . . digital cameras
H011 . . . . . film cameras
. . . . . *<film cameras by film size>*
H012 . . . . . 35 mm cameras
H013 . . . . . medium format cameras
H014 . . . . . miniature cameras
H015 . . . . . instant picture cameras
H016 . . . . . plate cameras
. . . . . *<cameras by viewing method>*
H017 . . . . . reflex cameras
H018 . . . . . single lens reflex cameras
H019 . . . . . twin lens reflex cameras
H020 . . . . . viewfinder cameras
H021 . . . . . special-purpose cameras
H022 . . . . . stereo cameras
H023 . . . . . underwater cameras
H024 . . . . . high pressure underwater cameras

(people)
. people (A100)
H025 . . models (people)
H026 . . photographers
(properties)
. properties (A200)
. . physical properties (A202)
. . . optical properties (A203)
. . . . colour (A204)
. . . . contrast (A205)
. . . . luminance (A206)
. . . . luminosity (A207)

P000 physics
(objects)
. equipment (A100)
P001 . . physics equipment
P002 . . . optical instruments
P003 . . . . light meters
P004 . . . . microscopes
(people)
. people (A100)
P005 . . physicists
(properties)
. properties (A200)
. . physical properties (A202)

Рисунок 7 — Классификационный указатель тезауруса, организованный по фасетам внутри тематических полей

На рисунке 7 также проиллюстрировано использование обозначений для ссылки на термин в классификационном указателе. В этом примере термины имеют только одно обозначение, соответствующее их основному месту. Когда термин повторяется в другом месте, он сопровождается своим обозначением (в скобках), которое помогает пользователю найти основное место термина. Некоторые тезаурусы позволяют терминам иметь дополнительные обозначения, так чтобы каждое повторение могло быть легко локализовано.

Выбор основного местоположения терминов является обычным методом сокращения объема печатных тезаурусов, что, однако, при полном изучении понятия приводит к необходимости осуществления дополнительных шагов. Выдача на компьютере может сделать процесс просмотра много легче, если пользователи могут разворачивать и сворачивать иерархию, выбирая только те детали, которые они хотят видеть.

12.2.5.3 Расширенный классификационный указатель

К типу указателя, который приведен на рисунке 7, чтобы представить термин во всех его локализациях, могут быть добавлены все данные, описанные в 12.2.2. Хотя на рисунке 8 это и не осуществлено в полной мере, он иллюстрирует включение дополнительной информации — лексических примечаний, аскрипторов, ассоциативных терминов, но не иноязычных эквивалентов.

Примечание — Сравните, например, underwater cameras (подводные камеры) на H023.

Включение дополнительных данных увеличивает объем. Для противодействия этому на рисунке 8 принято иное соглашение относительно повторения иерархий. Сравнивая рисунок 8 с рисунком 7, можно увидеть, что упоминание «pressure» и «temperature» в разделе рисунка 8, посвященного «подводному плаванию», в отличие от соответствующего места рисунка 7 не включает вышестоящих над этими терминами терминов «physical properties» и «properties». Чтобы их найти, пользователь должен пройти по связям, указанным метками A208 и A209, где показана полная иерархия. Кроме экономии места этот подход позволяет избежать создания ложного впечатления о содержании широких терминов, таких как «свойства» или «физические свойства», когда они встречаются во вторичных локализациях.

В заключение, обзор преимуществ указателей, приведенных на рисунках 6 — 8 следует отметить, что ни один из них не может быть рекомендован в качестве наилучшего во всех случаях. Каждый из подходов имеет преимущества и недостатки, оценка важности которых меняется в зависимости от контекста использования. Возможны и другие варианты: упорядочение по понятийным группам (см. 15.2.18). Однако необходимо, чтобы в дополнение к обычному алфавитному указателю пользователям была доступна, по крайней мере, одна форма систематического, иерархического или классификационного указателя (см. 12.3.4 относительно вопросов многоязычных тезаурусов).

A000 general terms

<*objects*>
A001 . equipment
. . <*equipment by portability*>
A002 . . . fixed equipment
A003 . . . portable equipment
. . <*equipment by power source*>
A004 . . . electrically-powered equipment
A005 . . . human-powered equipment
. . <*equipment by application*>
. . . diving equipment [D001]
. . . photographic equipment [H001]
. . . physics equipment [P001]

<*people*>
A100 . people
. . <*people by age*>
A101 . . . infants
A102 . . . children
A103 . . . adults
. . <*people by profession*>
. . . divers [D008]
. . . models (people) [H025]
. . . photographers [H026]
. . . physicists [P005]

<*properties*>
A200 . properties
A201 . . chemical properties
A202 . . physical properties
A203 . . . optical properties
A204 . . . . colour
A205 . . . . contrast
A206 . . . . luminance
RT: light meters [P003]
A207 . . . . luminosity
A208 . . . pressure
A209 . . . temperature

<*fields of work*>
A300 . fields of work
. . diving [D000]
. . photography [H000]
. . physics [P000]

D000 diving

<*objects*>
D001 . diving equipment
BT: equipment [A001]
D002 . . aqualungs
D003 . . diving suits
D004 . . . dry suits
D005 . . . wet suits
D006 . . face masks
D007 . . swimming fins
. . underwater cameras [H023]

<*people*>
D008 . divers
BT: people [A100]

<*properties*>
. pressure [A208]
. temperature [A209]

H000 photography

<*objects*>
H001 . photographic equipment
BT: equipment [A001]
H002 . . camera accessories
H003 . . . flash guns
. . . light meters [P003]
H004 . . . tripods
H005 . . cameras and camera components

photography (cont.)

<*objects*>(cont.)
. photographic equipment*(cont.)*
. . cameras and camera components *(cont.)*
H006 . . . camera components
H007 . . . . camera lenses
H008 . . . . camera viewfinders
H009 . . . cameras
. . . . <*cameras by imaging technique*>
H010 . . . . . digital cameras
UF+: *underwater digital cameras*
H011 . . . . . film cameras
. . . . . <*film cameras by film size*>
H012 . . . . . 35 mm cameras
H013 . . . . . medium format cameras
H014 . . . . . miniature cameras
H015 . . . . . instant picture cameras
SN: Cameras which produce a finished print directly
UF: *Polaroid® cameras*
H016 . . . . . plate cameras
. . . . <cameras by viewing method>
H017 . . . . . reflex cameras
SN: Cameras in which the image is reflected on to a glass screen for composing and focusing
H018 . . . . . single lens reflex cameras
UF: *SLR cameras*
H019 . . . . . twin lens reflex cameras
H020 . . . . . viewfinder cameras
H021 . . . . special-purpose cameras
H022 . . . . . stereo cameras
H023 . . . . . underwater cameras
SN: cameras for taking pictures under water
UF+: *underwater digital cameras*
BT: diving equipment [D001]
H024 . . . . . high pressure underwater cameras

<*people*>
. people [A100]
H025 . . models (people)
H026 . . photographers

<properties>
. optical properties [A203]
. . colour [A204]
. . contrast [A205]
. . luminance [A206]
. . luminosity [A207]

P000 physics

<*objects*>
P001 . physics equipment
BT: equipment [A100]
P002 . optical instruments
P003 . . light meters
UF: *exposure meters*
RT: luminance [A206]
P004 . . microscopes

<*people*>
P005 . physicists
BT: people [A100]

<*properties*>
. physical properties [A202]
. . optical properties [A203]
. . . colour [A204]
. . . contrast [A205]
. . . luminance [A206]
. . . luminosity [A207]
. . pressure [A208]
. . temperature [A209]

Рисунок 8 — Классификационный указатель тезауруса, организованный по тематическим полям, с дополнительной информацией — отношениями SN, UF, RT

12.2.6 Отображение полииерархических отношений

12.2.6.1 Стандартное обращение с полииерархическими отношениями (см. 10.2.5) показано на рисунке 5, где термин «underwater cameras» имеет два вышестоящих термина «special purpose cameras» и «diving equipment». Следовательно, «underwater cameras» появляется в двух позициях фасета

оборудования на рисунке 6. Если имеются еще и нижестоящие термины, то они также должны быть указаны в обеих позициях.

12.2.6.2 В классификационных и иерархических указателях печатного тезауруса, объем которых строго ограничен, ввиду того, что для полииерархических терминов требуется много места, может быть принято иное решение. А именно, термину разрешается иметь глубокую иерархию нижестоящих терминов, и, кроме того, длинное лексическое примечание, несколько синонимичных аскрипторов и ассоциированных терминов. Как показано на рисунках 6 и 7, для такого термина можно выбрать одно основное местоположение, а в других позициях показывать сокращенную информацию. Например вариант с термином «underwater cameras» на рисунке 8.

12.2.6.3 Компьютерная выдача должна позволять пользователям просмотр всех отношений полииерархического термина при каждом появлении.

12.3 Представление и оформление многоязычных тезаурусов

Раздел 12.3 опущен. В Российской Федерации вопросы многоязычных тезаурусов могут регулироваться по [11].

12.4 Вопросы языков и кодировки символов

Раздел 12.4 опущен. В Российской Федерации вопросы многоязычных тезаурусов могут регулироваться по [11].

13 Управление разработкой и ведением тезауруса

13.1 Проектирование тезауруса

13.1.1 Определение целей

Разработка тезауруса — это трудоемкая сложная работа, предполагающая долгосрочные обязательства, поэтому необходимо обосновать целесообразность такой разработки. Следовательно, прежде чем будут затрачены значительные средства и время, должны быть четко определены цели работы и даны ответы на следующие вопросы:

a) кто и для чего будет использовать тезаурус;

b) будет ли разработчик ограничен условиями существующего программного обеспечения, с которым он должен работать;

c) насколько глубокими должны быть знания пользователя о предметной области тезауруса и в вопросах использования тезаурусов.

Различные категории пользователей (руководители верхнего уровня, персонал поддержки информационных технологий или предполагаемые конечные пользователи) предполагают разные варианты того, каким тезаурус будет и, как он будет работать. В качестве ключевого шага к обеспечению успешной командной работы, положение о целях создания тезауруса должно быть согласовано со всеми заинтересованными сторонами.

13.1.2 Определение параметров тезауруса

Когда будут определены цели, назначения и круг пользователей, должна быть определена сфера охвата предметной области и стиль тезауруса.

Должны быть определены следующие свойства:

a) будет ли тезаурус использоваться в печатной или электронной форме, или в обеих формах;

b) с помощью каких указателей его следует предъявлять пользователю;

c) требуются ли дополнительные форматы для включения тезауруса в другие системы, такие как системы поиска и индексирования;

d) какой формат необходим для внесения изменений и как часто это потребуется делать.

Должно быть принято решение о факультативных свойствах, например, решение о том, следует ли различать разные типы иерархических отношений по 10.2.1, вводить ли какие-либо пользовательские отношения (см. 10.4), показывать ли в выходных формах метки узлов и важно ли сохранять данные определений, источников и дат работы с тезаурусом.

На более детальном уровне должны быть проработаны соглашения относительно использования единственного/множественного числа существительных, правописания и заглавных букв. Может оказаться полезным рассмотреть практические ограничения, такие, как максимальная допустимая длина

термина, круг необходимых знаков диакритики и специальных символов, и требуется ли ограничить число уровней иерархии.

13.1.3 Определение инвентаря ресурсов

13.1.3.1 Общие положения

На этапе планирования работ по тезаурусу должны быть определены ключевые ресурсы — персонал, финансирование, программные средства и словарные ресурсы. Если тезаурус должен быть многоязычным, в бюджет должны быть заложены дополнительные накладные расходы на общение, а также стоимость специализированных программных средств, которые могут понадобиться.

13.1.3.2 Человеческие ресурсы

Человеческие ресурсы должны включать:

a) руководителя на верхнем уровне менеджмента и, желательно, общее корпоративное соглашение о проекте тезауруса;

b) интерес и поддержку со стороны пользователей, особенно экспертов по терминологии в ряде разделов предметной области;

c) информационных работников, разбирающихся в предметной области и обладающих способностью, а еще лучше опытом работы с тезаурусами, которые могли бы принять на себя основную нагрузку в процессе создания тезауруса;

d) поддержку со стороны профессионалов в информационных технологиях, которые способны помочь в установке и адаптации системы управления тезаурусом, в программировании для интегрирования выходных форм тезауруса в системы, где он будет применяться, и в ежедневном обслуживании нормальной работы устройств на рабочем месте разработчиков.

Если в работе будет участвовать более одного редактора, должно быть проведено такое тщательное планирование работ, которое исключило бы возможность одновременного конфликтующего ввода данных в одно и то же место тезауруса.

Если создается многоязычный тезаурус, то персонал разработчиков и пользователей в идеале должен обладать следующими ниже перечисленными свойствами:

- хорошим пониманием каждого из используемых естественных языков;
- хорошим знанием предметной области;
- хорошим пониманием различия между переводом и установлением терминологических эквивалентов для целей информационного поиска.

Такие специалисты редки и обычно разбросаны по разным странам. Поэтому важно поддерживать постоянное интенсивное и эффективное общение членов коллектива разработчиков. Должны быть достигнуты соглашения по процедурам и форматам безошибочного обмена структурированными данными.

13.1.3.3 Словарные ресурсы

До начала построения тезауруса должны быть изучены имеющиеся словарные ресурсы. Могут оказаться полезными все следующее ресурсы:

a) существующие тезаурусы и классификационные системы с предметным охватом, полностью или частично совпадающим с требуемым. Может оказаться экономически выгодным просто адаптировать имеющийся тезаурус без каких-либо изменений. Однако более частотной является ситуация, когда в зависимости от авторских прав из существующих словарей удается заимствовать и/или адаптировать лишь определенную часть. Существующие разработки надо рассматривать как источник решений по структуре и составу терминов;

b) коллекции терминов и часто задаваемых вопросов, которые коллеги смогли собрать в ходе своей работы;

c) указатели содержания существующих баз данных или другие подобные справочные средства, указывающие точки входа в зависимости от тематики материалов. Если существующие близкие по тематике базы данных заиндексированы по контрольному словарю, может представиться возможность получить сведения обо всех использованных индексных терминах и о частоте их использования;

d) буфер запросов подходящего сайта, показывающий наиболее часто используемые поисковые термины;

e) стандартные справочные средства, такие как словари, терминологии, номенклатуры и энциклопедии полезны скорее для подтверждения значений терминов и обоснования иерархической структуры, чем для выбора терминов.

13.1.4 Распределение ответственности

На ранних стадиях разработки должен быть назначен редактор, который возьмет на себя ответственность за ведение проекта на всех его этапах, возможно включающее как составление первоначальной версии, так и ведение тезауруса в дальнейшем. В зависимости от объема работ в помощь редактору может быть образована редакционная коллегия, а возможно и коллегия рецензентов. Когда редактирование тезауруса соединено с анализом эффективности предметного индексирования по тезаурусу информационных ресурсов, это дает определенные преимущества.

13.1.5 Выбор программного обеспечения

Хотя тезаурус может быть создан без какого-либо программного обеспечения, такой ход событий не может быть рекомендован. Повышение эффективности разработки и возможность предотвращения канцелярских ошибок лежит на пути использования специально разработанной системы управления тезаурусом.

Для выбора программного обеспечения должен быть составлен перечень требований на основе положений раздела 14 и других соображений. Следует оценить все продукты, имеющиеся на рынке. Если ни один из них не обладает всеми требуемыми свойствами, может оказаться необходимым определенный компромисс. Если же компромисс невозможен, следует адаптировать к своим требованиям один из существующих продуктов, либо разработать новое программное обеспечение.

13.2 Ранняя стадия создания

13.2.1 Как и когда начинать

Поскольку тезаурус создается для поддержки индексирования и поиска в базе данных или другом ресурсе, тезаурус должен быть создан до начала наполнения базы данных, чтобы обеспечить систематичность индексирования. Однако до построения значительной части базы данных бывает трудно решить, какие термины необходимы в тезаурусе. Для того чтобы исключить последующее переиндексирование, хотя бы часть тезауруса должна быть готова заранее. Однако следует предусмотреть необходимость дальнейшего продолжения работы над тезаурусом.

Существуют программные средства, позволяющие проанализировать совокупность текстов в требуемой предметной области и выявить значимые слова и фразы с их частотой использования. Некоторые программные пакеты имеют возможность скомпоновать выявленные термины в структурированный формат. Хотя автоматически созданный тезаурус скорее всего не будет соответствовать качеству интеллектуально созданного тезауруса, но как источник идей его можно использовать для ускорения интеллектуального процесса.

13.2.2 Сбор терминов и понятий

Как отмечено в 5.1, термины в тезаурусе это только представители понятий, на которых основывается тезаурус. При построении тезауруса конечная цель состоит в сборе и организации понятий, но на практике это достигается сбором терминов и упорядочением их так, чтобы они адекватно представляли требуемые понятия.

Предпочтительные источники терминов перечислены в 13.1.4, перечисления а) — d). Необходимо иметь полный набор литературы, освещающей центральные вопросы разрабатываемой области знания. Как описано в 13.2.1, для выявления ключевых слов и фраз в текстах может быть использована автоматическая технология. Поскольку сравнительно просто исследовать материал, который уже введен в систему, необходимо, как можно больше внимания уделять выяснению того, что именно потребуется искать пользователям, например, анализировать списки запросов, которые были зарегистрированы справочной службой. При выработке предложений по включению в тезаурус понятий и терминов следует учитывать и мнение самих пользователей.

На начальной стадии нет необходимости принимать решения о правильном представлении понятий. Первоначальная задача состоит в собирании имеющихся терминов с указанием их источника и частотности.

13.2.3 Анализ терминов

Еще до ввода в тезаурус собранные термины должны быть рассортированы в систематическом порядке. Если система ведения тезауруса не обеспечивает поддержку этой задачи, то для первоначальной сортировки могут быть использованы стандартные программы персонального компьютера, например, электронные таблицы. При этом, если термины поучены из машиночитаемого источника, то этот процесс выполняется легко и быстро. Когда собранные термины будут записаны в электронных таблицах с их источником и частотностью, следует их распределить по предметным областям и фасетам (см. раздел 11) так, чтобы сходные термины со всеми их синонимами и вариантами оказались рядом.

Хотя классификационная и фасетная структура, используемая на этой стадии, может быть рудиментарной и не определяет возможную структуру тезауруса, она помогает выработке представления о такой структуре. Главным результатом этой аналитической стадии должен быть список или набор нескольких списков терминов, сгруппированных по темам или фасетам с указанием для каждого термина источника и частотности. Вооруженный этими данным, редактор имеет возможность начать систематическое построение тезауруса.

13.3 Построение тезауруса

13.3.1 Основные положения

Термины должны обрабатываться по тематическим/фасетным группам или по иерархиям. Группы синонимов и почти синонимов должны обрабатываться совместно, между ними должны быть установлены отношения эквивалентности. На этой стадии следует аккуратно выбирать термины и их иерархические отношения, принимая во внимание возможное дублирование, пересечение, пропуск терминов и необходимую степень подробности. В этом процессе могут быть выработаны и введены лексические примечания к терминам. Лучше всего, если запоминание даты ввода и формирование идентификатора термина будет осуществляться автоматически с помощью программного обеспечения, и не потребуется вводить их вручную с клавиатуры.

Если при предварительном сборе материала окажется, что существует тезаурус, определенная часть которого может быть заимствована, открывается возможность ввести ее непосредственно в используемую систему ведения тезауруса. Это сократит время, особенно если программное обеспечение позволяет ввод структурированных данных с лексическими примечаниями и отношениями терминов. Но при этом необходимо получить разрешение от правообладателей.

В некоторые программные продукты можно непосредственно ввести неструктурированные списки терминов. Это сохраняет не так уж много времени, поскольку необходимо еще вводить отношения и проводить тщательную проверку правописания, которое может быть случайно нарушено.

13.3.2 Последовательность построения тезауруса

При работе с иерархиями возникает вопрос, строить ли иерархию сверху вниз или снизу вверх. Необходимо иметь обзор верхних уровней, что обеспечит надлежаще сбалансированное тематическое покрытие тезауруса. Использование этого обзора помогает поместить предложения приглашенных специалистов и пользователей в контекст общей структуры.

С другой стороны, работа на более дробных уровнях имеет тенденцию вскрывать аномалии, способные повлиять на структуру верхних уровней. Наличие множества понятий нижних уровней иерархии, которые не находят подходящего размещения, говорит о необходимости развития соответствующей надструктуры.

Полезны обе стратегии построения и их следует применять комбинированно. Сначала до работы на нижних уровнях нужно подготовить заготовку структуры верхнего уровня. По мере продвижения работы на нижних уровнях заготовка верхних уровней должна изменяться в соответствии с возникающими потребностями.

Целесообразно работать с группами иерархически связанных терминов, вводя их совместно с отношениями эквивалентности и иерархии. На этой стадии отношения ассоциации не могут быть введены столь же легко, поскольку термины на другом конце ассоциаций обычно располагаются в другой иерархии, которую еще только предстоит разработать. Ассоциативные отношения обычно вводят на более поздней стадии, когда в системе представлен основной массив терминов.

При любой последовательности работ следует проводить контроль качества в таких аспектах как полнота, дублирование, правописание, корректный выбор отношений и т. п.

13.3.3 Построение многоязычного тезауруса

При построении многоязычного тезауруса могут применяться три подхода. Выбор подхода зависит от степени доступности человеческих, временных и финансовых ресурсов. Следует также принимать во внимание контекст предполагаемого использования, например, существует ли одноязычный тезаурус в рассматриваемой области, и использовался ли он для индексирования документов, к которым надо получить доступ.

а) Перевод одноязычного тезауруса

Перевод является распространенным и сравнительно недорогим путем построения многоязычного тезауруса. Он особенно выгоден, когда документы поискового массива уже были заиндексированы по одноязычному тезаурусу. Однако, если был принят этот подход, язык исходного одноязычного тезауруса приобретает статус доминирующего языка. Процесс перевода порождает в переводном языке

множество версий терминов и структур отношений, которые могут не соответствовать культурным и концептуальным ожиданиям коллективов пользователей в языке перевода.

b) Слияние нескольких различных одноязычных тезаурусов.

Слияние различных одноязычных тезаурусов является сложным процессом. Следует уважать целостность концептуальных и лингвистических систем и согласовывать их так, чтобы каждый язык мог поочередно выступать в роли исходного языка. Однако, поскольку различные тезаурусы обычно демонстрируют разные уровни специфичности и предкоординации, этот подход является наиболее трудным с точки зрения перспектив менеджмента.

c) Одновременное построение различных языковых версий тезауруса.

Этот третий подход дает возможность всем языкам в тезаурусе иметь равный статус. Каждый язык попеременно становится исходным и переводным и может воздействовать на установление иерархических и ассоциативных отношений. Работа должна быть организована таким образом, чтобы терминологическое и структурное согласование, требуемое для поддержания межъязыковых соответствий, делалось не всегда на одном и том же языке. Чтобы обеспечить успешный результат, с самого начала надо аккуратно координировать рабочую группу.

13.3.4 Привлечение экспертов и заинтересованных лиц

Следует воспользоваться следующими возможностями для привлечения к работе на разных стадиях других лиц.

a) Специальную терминологию иногда трудно охватить, и она может иметь неожиданные оттенки смысла. Но тот, кто хорошо знает специальную область, может легко проверить, является ли выбранная терминология корректной и правильно ли установлены отношения. Если термины используются неприемлемым для пользователей-экспертов образом, то тезаурусу нелегко будет получить признание. Следовательно, было бы правильно привлечь экспертов к проверке в ходе разработки тезауруса. Однако, как правило, термины имеют более чем одно значение, и значение, удовлетворяющее одного эксперта, может не получить одобрения другого. Более того, поскольку понятие, наиболее полезное для поиска информации, может иметь значение слегка более широкое или более узкое, чем термин, представляющий его, некоторые эксперты не смогут смириться с подобной неточностью. Привлекаемым экспертам должна быть хорошо поставлена задача, разъяснен порядок работы и условности контрольного словаря (см. раздел 4). Разработчик тезауруса должен быть готов направлять экспертов и помогать по формированию предложений.

b) Одному человеку трудно с первой же попытки принять все правильные решения. Большая часть достигается в ходе обсуждения трудных терминов с редакционной коллегией с некоторой долей побочных соображений. Если редакционная коллегия состоит из одного главного редактора, может быть полезным назначить комитет или группу специалистов для проверки разделов работы в процессе выполнения или для оказания консультаций по мере необходимости. Желательно, чтобы этому комитету редакционной проверки были бы представлены на одобрение все элементы тезауруса, прежде чем он будет издан.

c) Привлечение к делам время от времени старших руководителей и других заинтересованных лиц может способствовать дальнейшему финансированию, принятию и внедрению конечного продукта.

d) Привлечение пользователей может сильно увеличить закупку продукта. Люди, участвовавшие в работе, в большей степени склонны применять ее на деле.

e) Привлечение с самого начала экспертов на всех языках многоязычного тезауруса значительно улучшат перспективы его использования пользователями разных языковых сообществ.

13.4 Предисловие к тезаурусу

Все тезаурусы должны содержать подробное предисловие, которое ясно отражает цели тезауруса, а также зачем, как, кем тезаурус разработан. Должны быть освещены следующие вопросы:

a) охватываемая предметная область (области) с указанием ядерной зоны и периферийных зон;

b) языки, на которых представлен тезаурус;

c) подтверждение, какому национальному или международному стандарту соответствует разработка;

d) смысл всех соглашений, сокращений и пунктуационных знаков, используемых нестандартным способом;

e) общее число терминов с отдельным указанием дескрипторов и аскрипторов;

f) правила выбора понятий, включая выбор сложных понятий;

g) принятые правила выбора формы дескрипторов, включая ссылки на пособия, которым следовали разработчики, и правила разрешения их противоречий;

h) последовательность расположения и желательно указание, какому национальному или международному стандарту она соответствует;

i) указание политики обновлений, включая частоту, сроки и порядок обновлений, а также наименование и адрес ответственного органа, которому нужно направлять замечания и предложения;

j) ссылки и рекомендации в адрес источников, использованных для составления и проверки тезауруса.

Эти пункты по возможности нужно проиллюстрировать примерами и определениями технических терминов, использованных в предисловии.

Если тезаурус предназначен для какой-либо отдельной системы или программы, предисловие должно содержать указание о том, как тезаурус будет действовать в данном контексте.

13.5 Распространение тезауруса

13.5.1 Интеграция с электронной системой

Должны быть предусмотрены потребности людей, использующих тезаурус как для индексирования, так и для поиска. Удобнее всего, когда тезаурус со всеми его терминами, отношениями, лексическими примечаниями полностью интегрирован с функциями индексирования, поиска и просмотра, как указано в 16.3.

Некоторые системы индексирования и/или поиска имеют встроенный модуль ведения тезауруса. Если именно он использовался для разработки тезауруса, то не требуется никакой импорт/экспорт, и не нужна никакая специальная работа для этого. Однако, если интегрированная система не имеет механизма экспорта тезауруса в стандартный формат, могут возникнуть трудности при необходимости перейти на другую систему или сделать такой же тезаурус для другого применения.

13.5.2 Электронная публикация

Тезаурус может быть представлен как отдельный продукт, не связанный с каким-либо приложением. Электронный тезаурус можно распространять на CD-ROM, на сайте в Интернете или в локальной сети. Соответствующие функции просмотра описаны в 16.3.

Даже если тезаурус опубликован в сети без привязки к специальному приложению, его можно использовать для консультаций в реальном времени при проведении поиска в той же сети. Чтобы упростить эту задачу, нужно тезаурус представить в формате XML доступным в качестве «namespace».

В XML некоторые символы (а именно < > & ' «) имеют ограничения на употребление. По возможности их не следует использовать в тезаурусе (см. 6.4.2). Если же они необходимы пользователям (как в случаях «Boyle's Law, women's right»), то система ведения тезауруса должна их кодировать в выходной форме XML так, как изложено в стандарте XML [54], [55].

13.5.3 Ограниченное распространение печатных копий

Тезаурус можно предоставлять для ограниченного распространения в печатных копиях при следующих условиях:

a) полноформатный вывод с печатающего устройства системы ведения слишком дорог;

b) большинство пользователей предпочитают электронную версию;

c) тезаурус предназначен исключительно для собственного использования;

d) редакционная коллегия считает распечатку удобной для визуализации и подготовки замечаний.

В действительности такая функция является просто расширением обычной практики распечатки отчетов о состоянии тезауруса по мере надобности. Она обладает гибкостью дизайна выходных форм в соответствии с конкретными потребностями, например, включать и исключать служебные поля, выпускать подмножества тезауруса для отдельных групп пользователей, выбирать различные стили оформления и т. п. Такие отчеты можно рассылать пользователям в электронной форме, оставляя вопрос о копировании на решение пользователю.

13.5.4 Обычное издание

Хотя обычное издание уже в значительной мере уступило место электронному издательству, оно может оказаться полезным, если того требуют пользователи. Печатная форма имеет некоторые преимущества перед электронной, особенно в тех случаях, когда доступ к компьютеру ограничен. Кроме того, она может привлечь внимание к тезаурусу новой важной аудитории.

Когда организация принимает решение разработать и издать новый тезаурус, об этих планах должно быть объявлено в профессиональном журнале по соответствующей сфере деятельности, а также через соответствующие электронные дискуссионные форумы.

13.5.5 Депонирование в обменном фонде

Копия первого и всех последующих изданий опубликованного тезауруса должна быть депонирована в соответствующей национальной депозитарной библиотеке. Если тезаурус составлен на английском языке или включает английский язык в качестве одного из своих языков, копия тезауруса также должна быть депонирована в Североамериканском обменном фонде в Торонто[1)].

13.5.6 Каталоги тезаурусов в Интернете

Списки тезаурусов размещены на некоторых сайтах в Интернете. Некоторые из этих сайтов объявляют о разработке тезауруса до его публикации. Кроме того Инициатива метаданных Дублинского ядра (DCMI) поддержала несколько проектов разработки реестров метаданных, включая «реестр терминологии», предназначенный в том числе для контрольных словарей [29]. При публикации или до того разработчик должен послать полные данные о тезаурусе, будь он электронным или печатным, администратору подходящих сайтов или регистрационных служб. Если же электронный тезаурус доступен через Интернет, такая служба может давать прямую гиперссылку на него.

13.6 Обновления

13.6.1 Общие положения

Работы по ведению тезауруса должны продолжаться в течение всего жизненного цикла продукта. Необходимо отслеживать изменения как в терминологии, так и в составе базы данных или другого ресурса, который обслуживается тезаурусом. Потребность модернизации тезауруса возникает в день его выпуска, если не раньше. Предложения по внесению изменений могут быть наиболее частыми в первые месяцы после его выпуска, поскольку в это время словарь подвергается первому серьезному испытанию. Редакционные процедуры должны быть готовы с самого начала.

Хотя внесение изменений важно для того, чтобы тезаурус удовлетворял возникающим требованиям, следует помнить, что каждое изменение нарушает качество ретроспективного поиска (т. е. поиска по материалам, заиндексированным по прежней версии тезауруса). Когда вводят новый термин, появляется возможность поиска в базе данных материалов с данного момента, но релевантные ресурсы, уже имеющиеся в массиве, не использовали данный термин при индексировании. Если система управления базой данных обеспечивает автоматическое или полуавтоматическое исправление ретроспективных записей, изменения в составе дескрипторов могут быть встроены в поисковый массив. Однако другие типы изменений могут в меньшей мере влиять на ретроспективные записи. Следовательно, необходим компромисс между удовлетворением новых требований и поддержанием стабильности.

Если тезаурус многоязычный, перед одобрением изменений должны быть рассмотрены их влияния на все языки тезауруса.

13.6.2 Сбор предложений

Должен быть установлен простой механизм подготовки предложений ото всех пользователей, особенно от индексаторов. Этот механизм может быть представлен формуляром, электронным или бумажным, который приглашает сотрудников вносить необходимые им новые термины и другие изменения с подробным обоснованием. Обоснование включает определения необходимых терминов, ссылки на их употребление в литературе, синонимы и альтернативные способы выражения, связи с терминами, присутствующими в тезаурусе и объяснение, почему это изменение необходимо.

Индексаторы часто обнаруживают необходимость в новом термине в процессе ввода документов в базу данных. Поскольку при этом может не хватать времени для консультации с редактором, предлагаемый термин заявляется как «кандидат в дескрипторы». Кандидаты становятся доступными для поиска, но не проходят проверку на пригодность для тезауруса. Редактор тезауруса должен регулярно проверять все предложенные кандидаты и принимать решение об их вводе в систему ведения тезауруса со статусом «кандидат». Когда кандидат будет одобрен, производится соответствующая корректировка его статуса. Неодобренные кандидаты должны быть исключены из системы или исправлены подходящим способом.

13.6.3 Процедура пересмотра

Периодически редактор тезауруса должен просматривать все полученные предложения и кандидаты в дескрипторы. Кроме того он должен отслеживать частотность дескрипторов в обслуживаемых

[1)] Subject Analysis Systems Collection (SAS, Коллекция систем тематического анализа) поддерживается группой The Inforum: Факультет информации Университета Торонто, 140 St. George Street, Toronto, Ontario M5S 3G6, Canada. Подробности см. на gost.e-ecolog.ru . Фонд SAS включен в каталог университета.

базах данных. Каждый, используемый слишком часто или слишком редко, рассматривается как кандидат на исключение из тезауруса или на какое-либо изменение.

Для ведения также используются запросы пользователей. Термины, использованные в неудачных поисках, должны быть изучены на предмет пригодности в качестве дескрипторов тезауруса. Слишком обильное использование термина в запросах может свидетельствовать о том, что его нужно разделить по разным понятиям, а неиспользование, что нужно исключить из тезауруса.

Многие современные поисковые системы предлагают свободный текстовый поиск ресурсов и сохраняют буфер транзакций, включающий термины, введенные пользователями. Такой буфер может служить как действенное руководство по предпочтениям и привычкам пользователей, указывающее, например, какой раздел предметной области требует более детальной разработки, и какие понятия чаще требуют поиска — простые или сложные. Иногда буфер показывает, что нужно ввести дополнительный аскриптор к уже существующему дескриптору, или даже заменить дескриптор тезауруса синонимом, который используется чаще. Буферы транзакций должны полностью использоваться при ведении тезауруса.

Должны быть рассмотрены все предложения, термины пользователей и кандидаты индексаторов. Тщательное расследование может показать, что предложенные понятия уже присутствуют в тезаурусе под другим именем. Предложения чаще всего требуют введения нового дескриптора, но при этом редко принимают во внимание другие возможности и необходимые дополнительные действия, такие как введение нового ссылочного термина, изменение лексического примечания, добавление иерархических или ассоциативных отношений и т. п. Могут также требовать неприемлемого увеличения уровней иерархии. Редактор должен рассмотреть предложение в полный и согласованный набор необходимых изменений, прежде чем дать согласие по предложению. Если тезаурус — многоязычный, предложение должно быть рассмотрено с соответствующей адаптацией во всех языках. В малых предприятиях все решения вынужден принимать один редактор. Однако следует консультироваться с пользователями, особенно если тезаурус используется в разных приложениях. Бюллетень со всеми поступившими предложениями следует разослать по пользователям и рецензентам, представляющим все основные приложения, с предложением дать свои замечания. Рецензенты должны регулярно собираться или поддерживать общение в электронном форуме. После полного изучения и оформления изменений, требуемых предложениями, редактор должен представить их коллегии рецензентов. Прежде чем принять окончательное решение следует изучить последствия изменений для всех приложений тезауруса.

13.6.4 Виды изменений

Хотя главное внимание при ведении тезауруса делается на добавлении новых терминов, не следует пренебрегать и другими видами изменений. Встречаются следующие виды изменений, перечисленных в пунктах от a) до n):

a) добавление нового понятия, представленного новым дескриптором;

b) добавление нового аскриптора в качестве входного термина;

c) изменение дескриптора или аскриптора;

d) изъятие дескриптора;

e) перевод дескриптора в статус аскриптора, синонима другого существующего дескриптора:

f) изъятие аскриптора;

g) передача статуса предпочтительного термина с дескриптора на один из синонимичных ему аскрипторов;

h) добавление или изменение лексического примечания, определения, исторического или редакторского примечания;

i) слияние двух понятий в одно;

j) разделение понятия на два или более, что иногда может сопровождаться преобразованием аскрипторов в дескрипторы;

k) изменение иерархической структуры;

l) перенос ветви иерархии с одного места в другое;

m) добавление или удаление ассоциативных отношений;

n) почти любая комбинация изменений по пунктам от a) до m), которая может иногда повлечь различные изменения в различных языках многоязычного тезауруса. Например, добавление аскриптора в одном языке может требовать, а может и не требовать добавления аскриптора в другом языке.

Важно чтобы критерии принятия редакционных решений сохранялись при анализе новых предложений и включали критерий удобства для пользователей. Особенно внимательно нужно относиться к исключению терминов, особенно если они использовались в прошлом. Вместо исключения дескрип-

тора его статус может быть понижен до аскриптора со ссылкой на соответствующий дескриптор и с отметкой в историческом примечании, до какого времени термин имел прежний статус. Соответствующий дескриптор также должен иметь в историческом примечании отметку об этом изменении.

При сложных изменениях историческое примечание может составлять редактор, но даты ввода и изменения данных должны записываться автоматически. Аналогичным образом идентификаторы новых терминов должны также присваиваться автоматически.

Изменения в многоязычном тезаурусе должны вводиться одновременно во всех языках.

13.6.5 Информация об изменениях

13.6.5.1 Общие положения

Механизм распространения сведений об изменениях должен быть продуман до издания тезауруса. Информация об изменении может быть сложной, особенно если атрибуты и отношения одного термина претерпевают несколько различных изменений. При прямой рассылке новых терминов пользователям будет трудно понять, какому месту в иерархии соответствуют новые термины, особенно если необходимо рассылать различные виды указателей. Предполагая, что простой список новых терминов достаточен для пользователей, можно прийти к ситуации, что новые термины не будут использоваться, поскольку они не будут полностью интегрированы в механизм просмотра тезауруса.

Простейший путь распространения информации об изменениях — это рассылка новой версии тезауруса. Если это делать электронными средствами, цена сводится к минимуму. Более того, это исключает обмен сведениями о характере изменений и гарантирует, что все пользователи будут иметь доступ к единой авторитетной версии словаря при выборе терминов. К тому же эта процедура позволяет сообщить сведения об изменениях одновременно персоналу и компьютерной системе.

13.6.5.2 Изменения, дружественные к читателю

Для конечных пользователей информация об изменениях может иметь форму алфавитного списка, дополненного классифицированным или категоризированным списком, в которых указаны все новые добавленные дескрипторы, измененные дескрипторы, новые входные термины. В этой информации должно быть приглашение к ответному сообщению. В зависимости от пожеланий пользователей список может иметь разную форму сложности. Некоторые пользователи могут иметь желание получать полную информацию о каждом новом или измененном термине и отношении. Электронные формы могут информировать об изменениях выделением их в тексте цветом или подчеркиванием. Всем пользователям должно быть предложено отмечать наиболее важные изменения, такие как введение и исключения терминов в их бумажном экземпляре тезауруса, если они им пользуются. Для печатного тезауруса это облегчается изданием его в форме разъемных листов, а информация об изменениях включает новые листы издания, подлежащие замене.

13.6.5.3 Изменения для компьютеров

Если копия тезауруса встроена в другую компьютерную систему, то простейшим путем будет полная замена его на новую версию, в которой уже сделаны все изменения. Однако это может быть неприемлемо, если в системе тезаурусные термины связаны с другими данными, такими как библиографические записи или отображения на другие словари. Для таких систем должны быть предоставлены как минимум отчеты или файлы изменений всех элементов, атрибутов и отношений (см. раздел 15). Для поддержки этих требований система ведения тезауруса должна быть способна записывать дату последних изменений всех элементов, атрибутов и отношений и по этой дате выдавать сведения об измененных данных. Альтернативная технология может предусматривать создание внешнего буферного файла, в котором записываются все произведенные добавления, исключения и изменения элементов, атрибутов и отношений. Для импортирующей системы некоторые типы изменений могут быть существенными, а другие — не существенными. По этой причине некоторые системы ведения тезауруса записывают тип изменения и само изменение отдельно. Это позволяет импортирующей системе отделять простые изменения (например, орфографии) от сложных (например, разделения одного понятия на два). Поскольку стандартная классификация изменений не существует, ее следует разрабатывать для каждого случая заново.

13.6.5.4 Следствия для редакционного хозяйства

Важность поля даты последних изменений уже подчеркивалась в связи с представлением изменений для компьютеров. Оно также иногда необходимо (наряду с датой ввода) в данных для визуального представления изменений. Эта дата должна вводиться в стандартном формате ГГГГ-ММ-ДД (например, 2004-03-17) в соответствии с [12]. В случае изменения написания термина также жизненно необходим идентификатор термина. (В отсутствие идентификатора компьютер не имеет способа уста-

новить, какой термин должен быть заменен на термин с новым написанием). Все упомянутые атрибуты должны записываться автоматически во избежание человеческих ошибок.

Другой жизненной необходимостью редакционных групп является список исключенных терминов. Однако не многие системы ведения тезауруса имеют механизм запоминания выдачи списка удаленных терминов. Поэтому редактор сам должен вести такой список, включающий термины и их идентификаторы.

14 Указания по программному обеспечению для ведения тезауруса

14.1 Общие положения

Программное обеспечение для ведения тезауруса должно быть способно управлять, импортировать и экспортировать данные согласно положениям раздела 15. Оно должно быть выбрано на основе требований имеющегося проекта (см. 13.1.5). Обычно от него требуется наличие характеристик и функций, описанных в 14.2 — 14.10, но могут быть исключения. Также иногда требуются дополнительные свойства. Например, полезно, если программное обеспечение обладает способностью одновременного ведения двух и более тезаурусов.

Примечание — В этом документе не идет речь о требованиях, обычно предъявляемых к программному обеспечению в целом, таких как хорошее документирование, обучение и другие формы поддержки, общая дружественность интерфейса к пользователю и приемлемая цена.

14.2 Ограничения по объему и символам

Программное обеспечение не должно накладывать следующих ограничений.

a) Не должно быть ограничений по числу терминов в словаре, которые бы препятствовали его расширению до необходимых размеров. То же самое относится к другим элементам — меткам узлов, примечаниям.

b) Желательно, чтобы не было ограничений на длину терминов, меток узлов, примечаний. Хотя довольно редко длина терминов бывает более 40 знаков, но в некоторых случаях может потребоваться 100 или более знаков.

c) Не должно быть ограничений на допускаемое число уровней иерархии, а также на число иерархических, ассоциативных и синонимических связей у дескрипторов.

d) Программное обеспечение должно работать со всеми символами Универсального набора знаков по ИСО/МЭК 1-646 [13] в текстах элементов тезауруса (включая термины, метки узлов и примечания).

e) Редактор должен иметь возможность свободно выбирать верхний или нижний регистр литер при необходимости.

f) Для многоязычных тезаурусов не должно быть ограничений на число языков тезауруса. Должна предусматриваться возможность задавать все текстовые элементы (такие как термины, метки узлов и примечания) на каждом из языков тезауруса.

14.3 Отношения между терминами и между понятиями

Программное обеспечение должно удовлетворять требованиям перечислений a) — i), приведенным ниже.

Примечание — Там, где в настоящем разделе указываются английские метки BT, NT, RT, USE, UF, SN, в других языках должны применяться соответствующие национальные метки (см. таблицу 2 в разделе 3).

a) Не может быть в тезаурусе совпадающих терминов. При вводе термина, совпадающего с уже существующим, он должен быть исключен или выдан редактору для изменения, объединения, добавления квалификатора или для другого способа исправления. Алгоритм сравнения для выявления дублей должен иметь возможность такой настройки, чтобы, например, типографские различия, такие как курсив или выделение заглавными буквами, не принимались в расчет при выявлении дублей.

b) Должны поддерживаться базисные отношения BT/NT, RT/RT, USE/UF, как это описано в разделах 8 и 10.

c) Должны поддерживаться отношения взаимности, описанные в разделах 8 и 10. Если понятие A имеет отношение BT с понятием B, то понятие B должно иметь отношение NT с понятием A и наоборот. Если понятие C имеет отношение RT с понятием D, то понятие D должно иметь также отношение RT с понятием C. Если термин E имеет отношение USE с термином F, то термин F должен иметь отношение

UF с термином F и наоборот. Желательно, чтобы программное обеспечение создавало обратное отношение автоматически в ответ на введение редактором исходного отношения. Как минимум программное обеспечение должно выдавать предупреждение о наличии невзаимных отношений.

d) Когда термин или понятие изменяется или исключается, то это изменение должно автоматически распространяться на все случаи, где этот термин или понятие связаны с другим термином или понятием любым отношением — BT, NT, RT, USE, UF. При исключении термина или понятия все связи к нему или от него должны быть изъяты. Однако если исключение термина, понятия или отношения оставляет какое-либо понятие без по крайней мере одного вышестоящего или без предпочтительного термина (дескриптора), должно быть выдано соответствующее предупреждение редактору.

e) Не должно быть ограничений на число отношений, которые имеют данное понятие или дескриптор. Так одно понятие может иметь любое число отношений любого типа — BT, NT, RT, а каждый дескриптор — любое число аскрипторов.

Примечание — Некоторые комбинации недопустимы (см. перечисления g) ниже).

f) Должна быть предусмотрена возможность вводить пользовательские взаимные отношения, например, для того, чтобы различать различные типы отношений BT/NT или различные типы ассоциативных отношений.

g) Проверка правильности данных должна предотвращать ввод недопустимых комбинаций отношений, приведенных ниже:

- если два термина или понятия уже связаны одним из базовых отношений, никакое другое базовое отношение между ними введено быть не может;
- если понятие A имеет отношение BT с понятием B, то никакое понятие, стоящее в иерархии BT над понятием B, не должно получать отношений BT, NT, RT, UF с понятием A;
- непредпочтительные термины (аскрипторы, т. е. термины с отношениями USE, USE+ к другим терминам) не могут иметь отношений BT, NT, RT, UF;
- в случае отношений USE+ программное обеспечение должно проверять, что это отношение является по крайней мере тернарным.[1)]

h) Не допускаются отношения терминов и понятий к самому себе, т. е. все базовые отношения нерефлексивны.

i) Для каждого понятия допускается иметь только один предпочтительный термин (дескриптор) в каждом из языков тезауруса.

14.4 Примечание к терминам и понятиям

Программное обеспечение должно поддерживать примечания следующим образом:

a) Должен поддерживаться ввод лексических примечаний к каждому понятию.

Примечание — Лексические примечания могут иметь произвольную длину.

b) Если примечание любого типа имеет ссылку на другой термин или понятие в словаре, желательно, чтобы программное обеспечение обеспечивало маркировку или гиперссылку к записи соответствующего термина или понятия. Программное обеспечение должно проверять действенность ссылки.

c) Должно поддерживаться формирование автоматических примечаний к терминам и понятиям, например, примечаний об изменениях и редактировании.

14.5 Коды и обозначения

Программное обеспечение должно поддерживать предметные категории, коды и обозначения следующим образом:

a) должна быть предусмотрена возможность присвоить по крайней мере один код, номер, или другое обозначение каждому термину, понятию, понятийной группе и ряду (см. 12.2.2 и 15.2.18). Желательно, чтобы поддерживалось присвоение нескольких обозначений.

b) Должна быть возможность присвоить по крайней мере одну предметную категорию каждому термину, понятию, понятийной группе и ряду (см. 12.2.2 и 15.2.18).

c) Должна быть возможность присвоить уникальный идентификатор каждому термину и каждому понятию. Желательно, чтобы присвоение идентификаторов было автоматическим при вводе каждого

[1)] Имеется по крайней мере два термина, на которые ссылается отношение USE+ (прим. переводчика).

нового термина и понятия и производилось таким образом, чтобы не появлялись дубли идентификаторов существующих или исключенных терминов и понятий.

d) Должна быть возможность вывести словарь в последовательности любого типа обозначений, кодов и идентификаторов.

14.6 Метки узлов

Программное обеспечение должно поддерживать метки узлов следующим образом:

a) метки узлов не рассматриваются как термины или понятия тезауруса (см. раздел 11, рисунок 4) и следовательно не связаны требованиями к отношениям п. 14.3. Более того, они не обязаны быть уникальными и не подлежат контролю на дублирование, описанному в п. 14.3, перечисление a);

b) программное обеспечение должно иметь средства показать на выдаче метку узла в правильном положении относительно родительских терминов и относительно терминов верхнего уровня, которые имеются у данного фасета или ряда.

14.7 Статус языков

Программное обеспечение должно позволять всем языкам многоязычного тезауруса иметь равный статус и избегать преимуществ одного языка по сравнению с другими. Например:

a) Число аскрипторов, приписанных какому-либо понятию в одном языке, не должно зависеть от числа аскрипторов в другом языке, и аскрипторы в различных языках не обязаны соответствовать друг другу.

b) Наличие лексического примечания к понятию в одном языке не должно требовать наличия соответствующего лексического примечания в другом языке.

14.8 Импорт/экспорт данных

Необходимо, чтобы была возможность экспорта и импорта данных таким образом, чтобы тезаурус не был ограничен тем программным обеспечением, в котором он был первоначально создан. Форматы импорта/экспорта обсуждаются в разделе 17. Должны быть предусмотрены следующие механизмы:

a) Полный вывод данных существующего словаря, включая термины, примечания, метки узлов, стандартные отношения терминов и понятий и все другие атрибуты терминов, понятий и меток узлов.

— При выводе должны сохраняться все упомянутые особенности, а также все используемые символы Универсального набора знаков.

b) Формирование отчетов и вывод словаря, включая все термины, примечания, обозначения и стандартные отношения между терминами и понятиями.

— Должна существовать возможность вывода подмножеств, определяемых редактором, например, таких как только непредпочтительные термины (аскрипторы) или только предпочтительные термины (дескрипторы) с их лексическими примечаниями и нижестоящими связями.

c) Пакетное редактирование.

— Должна существовать возможность редактирования или исключения пакетов записей одновременно, предпочтительно с использованием собственных возможностей программного обеспечения. Если же это невозможно, то должны использоваться функции базовой системы ведения баз данных или процедура вывода избранных записей, редактирования их вне тезауруса и повторного импорта в тезаурус. Эта техника может использоваться, например, для ввода одинаковых исторических справок или отношений к пакету терминов или, может быть, для отображения терминов на другой словарь. Если применяется этот метод, должна также производиться обычная проверка целостности данных.

d) Экспорт всех терминов, которые были изменены после некоторой даты, с их полными статьями или без них.

— Должна быть доступна возможность выбора только определенного типа изменений, например, только новые термины или термины на одном из языков, или включая все термины, в которых изменены какие-либо атрибуты и отношения. Также должна быть возможность получить список всех терминов, исключенных с определенной даты.

e) Вывод указателей тезауруса, либо на экран, либо на печать.

— Должна быть возможность выбора между различными последовательностями и формами указателей, из которых наиболее важен алфавитный указатель, но также настоятельно рекомендуется иерархический или классификационный указатель.

f) Когда родственные термины в алфавитном, иерархическом или классификационном указателе представляют в алфавитном порядке, может быть использовано соглашение о расположении «слово за словом», либо «буква за буквой». Такое же соглашение должно применяться ко всему тезаурусу. (Объяснение этих соглашений описано в [7]; см. пример ниже). Однако должна быть возможность нарушать эту последовательность в случае необходимости (см. раздел 11).

Пример —

Последовательность «слово за словом»	*Последовательность «буква за буквой»*
National insurance	***National insurance***
National parks	***Nationality***
Nationality	***National parks***

14.9 Редакторская навигация и поддержка

Программное обеспечение должно поддерживать следующие редакторские функции.

a) Редактор тезауруса должен иметь возможность получить все детали любого термина (и соответствующего понятия) путем прямого обращения к этому термину, предпочтительно, путем ввода некоторой части термина при том, что программное обеспечение выберет все термины, содержащие эту часть.

b) Редактор тезауруса должен иметь возможность получить все детали любого термина (и соответствующего понятия) путем просмотра связей термина/понятия. Желательно, чтобы функция просмотра позволяла редактору начинать со списка терминов верхнего уровня, но при отсутствии этой возможности редактор должен иметь возможность переходить от любого термина к записям любого связанного с ним термина отношениями NT, BT, RT, USE, UF.

c) Редакторский интерфейс должен позволять просмотр полного иерархического контекста термина или понятия одновременно с самим термином со всеми его примечаниями, кодами и отношениями, представленными для редактирования.

d) Должна быть возможность легкого перемещения понятия (вместе со всеми его нижестоящими всех уровней) из одной позиции в иерархическом дереве в другую, желательно с помощью перетаскивания курсором.

e) Редакторские возможности должны включать стандартные функции обработки слов, такие как добавление, изменение и исключение букв без необходимости повторного ввода с клавиатуры всего поля. Необходима функция проверки правописания.

f) Когда редактор предпринимает шаги к исключению термина или понятия, программное обеспечение должно просить подтверждения намерению исключить термин или понятие прежде, чем факт исключения будет совершен.

g) Установление нового отношения между двумя терминами или понятиями, уже представленными в тезаурусе, должно производиться путем навигации и выбора терминов, а не путем повторного набора термина, известного системе. Это необходимо для предотвращения ошибок и повышения эффективности работы.

h) Редактор должен иметь возможность легко переключаться с выдачи на одном языке к соответствующей выдаче на другом языке тезауруса.

14.10 Редакционная безопасность

Программное обеспечение должно предусматривать следующие меры предосторожности.

a) Редакционные изменения должны первоначально производиться в эталонной базе данных, из которой тезаурус периодически выгружается для текущих процессов индексирования или поиска ресурсов.

b) Если эталонную базу данных редактируют одновременно два лица, встроенный механизм должен предотвращать одновременный доступ для изменения одной и той же записи.

c) Проверка безопасности по паролю должна предотвращать введение редакционных правок посторонними лицами.

d) Желательно, чтобы программное обеспечение предусматривало разные уровни доступа, так чтобы можно было вводить предварительные изменения, которые вводятся в действие после одобрения.

e) Должна быть предусмотрена функция возврата к прежнему состоянию, которая позволяет каждому редактору отменить некоторое количество последних изменений.

f) Должен поддерживаться буфер, позволяющий восстановить базу данных по предыдущей версии.

14.11 Средства учета

Должна иметься возможность получить отчет о числе терминов с определенными характеристиками, в частности — общее число дескрипторов и аскрипторов.

Как описано в 13.6.3, должен вестись учет числа использований терминов для индексирования и поиска. Также программное обеспечение тезауруса должно иметь возможность импорта данных об использовании, если это не выполняется каким-либо другим средством. Для хранения этих данных может потребоваться одно или более пользовательских полей или атрибутов (см. 15.2.16).

15 Модель данных

15.1 Общие положения

Этот раздел моделирует структуру данных, которую можно принять за основу построения тезауруса, удовлетворяющего указанным выше требованиям ИСО 25964. Для многоязычных тезаурусов, однако, эта модель данных применима только в случае полной симметричности. Если структуры данных для разных языков различаются, то может понадобиться особая техника отображения, по крайней мере для частей тезауруса, где это различие имеет место. Техника отображения будет рассмотрена в [6].

Схема XML выведена из этой модели и может быть использована при передаче или приеме тезауруса в электронной форме. Настоящая модель представляет логическую структуру, которую необходимо поддерживать, но она не указывает, как должны быть реализованы данные в конкретных компьютерах.

Базовые рекомендации разделов с 5-го по 12-й достаточно просты и требуют только простой модели. Но реализация нескольких факультативных свойств потребует более сложной модели. Для обеспечения успешного обмена тезаурусными данными в большем диапазоне ситуаций важно, чтобы имелся формат, восприимчивый как к сложным, так и к упрощенным тезаурусам.

Модель, представленная здесь, включает полный набор возможностей, описанных в разделах с 5-го по 12-й. Некоторые элементы модели показаны как факультативные, и они не требуются при обмене тезаурусами, в которых эти элементы отсутствуют. Если программное обеспечение, импортирующее тезаурус, не поддерживает факультативные элементы, оно может их игнорировать, а также принимать подходящие значения по умолчанию (например, нуль) для тех факультативных элементов данных, которые оно поддерживает, но которые отсутствуют в принимаемом тезаурусе, при этом корректно интерпретируя остальные элементы.

Модель представлена сначала на рисунке 15 с помощью соглашений UML (Unified Modelling Language — Унифицированный язык моделирования) [14], [15], [16], но четко останавливаясь перед описанием поведения тезауруса и методов распределения, которые могут потребоваться в спецификации по установке программного обеспечения. UML достаточно выразителен и исключает неоднозначности, но поскольку некоторые пользователи могут быть незнакомы с этими соглашениями, в 15.3 дополнительно приведено альтернативное табличное представление. Элементы данных и атрибуты в диаграммах UML и в таблицах данных идентичны по содержанию, хотя некоторые детали отношений не удается ясно выразить в табличном представлении. В случаях сомнений следует обращаться к диаграммам.

UML содержит собственные специальные соглашения и термины для моделирования. В настоящем разделе термин «класс» используется в смысле UML для обозначения множества объектов, имеющих сходную структуру, поведение и отношения. Также UML придает специальный смысл терминам «строка», «булевский» и «агрегация». Дополнительные сведения об UML можно найти в библиографии по ссылке [16], где глава 5 — руководство по нотации UML, а раздел 4.5.4 детально описывает семантику.

Замечания по модели, имеющиеся в 15.2, относятся как к табличному представлению, так и к диаграммам на рисунке 15.

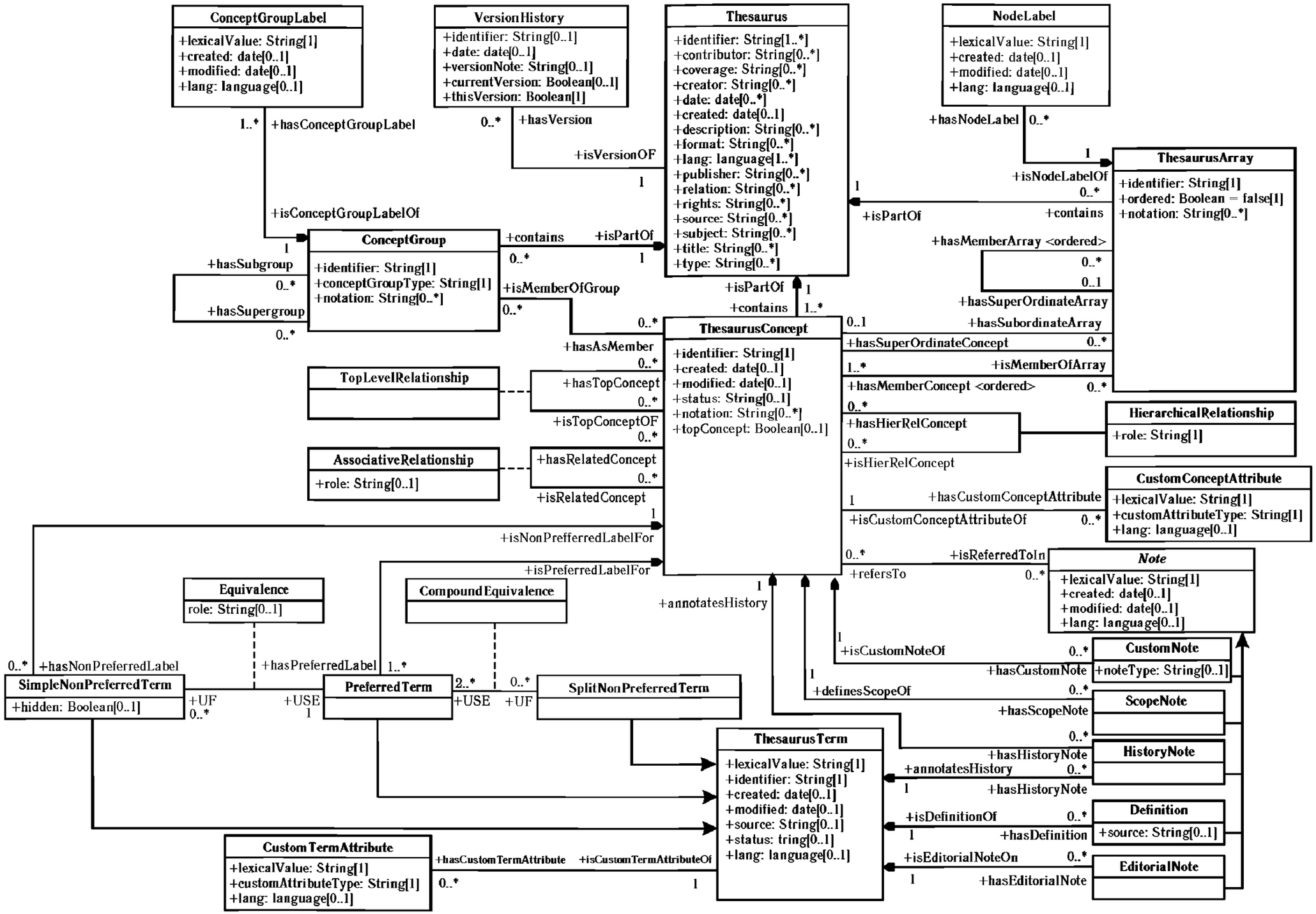


Рисунок 15 — Модель структуры данных построения тезауруса

15.2 Замечания по модели

15.2.1 Соглашения об именах

Имена всех атрибутов, включая некоторые элементы данных Дублинского ядра (см. ссылку [17]), которые появляются в модели и в ее схеме XML, следуют «соглашению о верблюдеНижнегоРегистра» (т. е. между словами не делается пробел, а каждое следующее слово в имени (кроме первого) начинается с заглавной буквы). Имена классов UML следуют «соглашению о ВерблюдеВерхнегоРегистра» (т.е. пробелы отсутствуют и каждое слово в имени начинается с заглавной буквы).

15.2.2 Символы для ассоциаций

Ромб на диаграммах обозначает «агрегацию», т. е. отношение обладания (has-a отношение). Этот символ используется для указания, что в соответствующей схеме XML определение объекта включено в родительский объект.

Символ треугольника обозначает «генерализацию», т. е. отношение вида к роду (is-a отношение), связывая обобщающий класс с его видовыми подклассами, которые наследуют все атрибуты родительского класса.

Обычная ассоциация (без ромба) обозначает отношение типа key/keyref в XML, где каждый экземпляр одного класса содержит ключ ссылки, который связывает его с экземпляром другого класса. Заметим, что схема XML в действительности не навязывает ограничений key/keyref. Объясняется это тем, что эти ограничения приводили бы к обесцениванию выходных файлов XML, которые содержали бы только фрагмент тезауруса, а не весь тезаурус целиком.

15.2.3 Понятие

Каждое понятие в тезаурусе представлено одним дескриптором в каждом языке, и некоторым количеством аскрипторов. Нотация, примечания и родовидовые отношения применяются к понятию в целом, а не к дескриптору как таковому. Каждому понятию может быть присвоен идентификатор. В некоторых системах понятие идентифицируется только самим дескриптором или идентификатором дескриптора, но это делать не целесообразно, поскольку написание терминов может меняться.

15.2.4 Идентификатор

Классы *Thesaurus, ThesaurusConcept, ThesaurusTerm, ConceptGroup* — все они имеют уникальные идентификаторы, которые используются в схеме XML. Идентификатор и значения дат для ThesaurusTerm существенны в отношении организации службы обновлений, поскольку, когда меняется написание термина, преемственность данных сохраняется благодаря сохранению идентификатора в цепи последовательных обновлений. Использование идентификаторов понятий строго рекомендуется для обеспечения совместимости в среде сетевых приложений.

15.2.5 Даты

Даты создания и изменения строго рекомендуется указывать, несмотря на то что соответствующие элементы данных в некоторых классах являются факультативными. Они должны быть даны в формате ГГГГ — ММ — ДД в соответствии с [12]. В классе **Thesaurus** имеется альтернатива — использовать простой атрибут «*date*», который будет указывать, когда была создана версия того, к чему этот атрибут привязан. Формат даты необходимо использовать тот же.

15.2.6 Версии и их история

Класс *VersionHistory* позволяет по желанию иметь в каждой копии запись о версиях или изданиях, которые были выпущены.

Хотя этот класс факультативен и может не понадобиться, если существует только одна версия, но крайне рекомендуется принять его, если версий больше, чем одна. Каждая версия должна быть идентифицирована идентификатором, или датой, или и тем и другим одновременно.

Атрибут *versionNote* используют для указания на природу версии, т. е. создана ли она как обновление, как фрагмент, как перевод, а также для разъяснения отношений к другим версиям.

Атрибут *curentVersion* — это булевский флажок (да/нет) для указания в каждой версии, является ли она рабочей или она заменена и отброшена. Рабочими могут быть одновременно несколько версий.

Атрибут *thisVersion* — это булевский флажок для указания того, к какой из перечисленных версий эта история прилагается.

Решить, какой из вариантов тезауруса является «версией», а какой должен рассматриваться как самостоятельная разработка, не всегда просто. В настоящем контексте управление версиями предполагает главным образом последовательное обновление исходного тезауруса. Таким образом:

а) обновления, сделанные или принятые первоначальным владельцем/создателем — новые понятия, термины, отношения, примечания, изменения в существующих частях тезауруса, а также специальные издания для различных целей — должны рассматриваться как новые версии данного тезауруса;

b) подмножества тезауруса, даже если они имеют подзаголовок «Сокращенная версия», более целесообразно рассматривать как «понятийную группу» (см. 15.2.18), если они основаны на действующей версии;

c) подобным образом языковые версии, при условии что они соответствуют исходной разработке и имеют общую структуру понятий (даже если они могут быть названы «Английская версия», «Французская версия») должны рассматриваться как принадлежащие той же самой версии тезауруса. С целью обмена данными они должны дифференцироваться при помощи языкового атрибута класса *Thesaurus*.

d) Всякая разработка, включающая изменения терминов, понятий или отношений, новые переводы, упрощения/сокращения отношений, если ее представляет кто-либо кроме первоначального владельца/создателя, должна идентифицироваться как отдельный тезаурус. Указывать на то, что это отдельный тезаурус могут изменения в атрибутах *creator, publisher, rights, title* класса *Thesaurus*. В этих случаях должен быть назначен отдельный идентификатор и дана связь к исходной работе в данных таких атрибутов как *relation, description, source*.

15.2.7 Статус

Status — факультативный атрибут в *ThesaurusConcept* и *ThesaurusTerm*, который указывает на то, является ли данный объект, например, утвержденным кандидатом, замененным или исключенным (см. 13.6.2).

15.2.8 Нотация

Notation — факультативный атрибут в *ThesaurusConcept, NodeLabel* и *ConceptGroup* (см. 12.1.3 и 12.2.5.2).

Если тезаурус использует выразительную нотацию, то, присоединяя ее к меткам узлов, при иерархическом представлении мы получаем возможность указывать узлы в надлежащем месте.

Если выразительная нотация отсутствует, то должны быть найдены другие средства, чтобы выводить метки узлов в правильной последовательности, такие как код сортировки, который скрыт от пользователей. В этом случае тот же самый атрибут требуется для вывода дескрипторов.

Система нотации в *ConceptGroup* вполне может отличаться от нотации в *ThesayrusConcept*, и каждая из этих систем может быть представлена без другой.

15.2.9 Источник

Source — факультативный атрибут в *Thesaurus* и *ThesaurusTerm,* который можно использовать для ссылки на источник сведений или на разработчиков словарной статьи.

15.2.10 Коды языков

Language — факультативный атрибут в *ThesaurusTerm, Note* и *NodeLable* и обязательный в *Thesaurus*, если он имеется в списке, то он должен быть представлен кодом alpha-2 из [18] или [19], эти коды при необходимости могут быть расширены дополнительными кодами, описанными в RFC 4646 [20] и перечисленными в регистре подметок (субтэгов) IANA [21] (см. 12.4.5).

15.2.11 Элементы Дублинского ядра

Класс *Thesaurus* имеет несколько атрибутов для указания метаданных Дублинского ядра [17] в формате [7]. Наряду с описанными выше атрибутами *identifier, date, source, language* сюда относятся contributor, *coverage, creator, description, format, publisher, relation, rights, subject, title u type*. Все они, кроме *language* и *identifier,* факультативны.

15.2.12 Множественность дескрипторов

Модель позволяет иметь в тезаурусе несколько дескрипторов для обозначения одного и того же понятия, что необходимо в случае многоязычных тезаурусов, где имеется по одному дескриптору на каждом языке. Для одноязычных тезаурусов такая множественность исключается.

15.2.13 Корневые (вершинные) понятия

В разделе 12 кратко упоминается о терминах верхнего уровня («Top Terms»), т. е. о терминах, представляющих понятия самого верхнего уровня в качестве факультативной составляющей алфавитного указателя и единого списка дескрипторов (см. 12.2.2 и 12.1.3), которая связывает понятие с вершиной иерархии, в которой находится понятие. Обеспечивает это класс *TopLevelRelationship*.

Рекомендуется иметь возможность распечатки редакторов со списком вершинных терминов (14.9 перечисление b). На это свойство указывает атрибут topConcept, помечающий понятия, которые находятся наверху иерархии, т. е. не имеют вышестоящих понятий. При импортировании тезауруса это облегчает возможность выбирать все вершинные понятия и строить на них иерархические деревья.

15.2.14 Примечания

В тезаурусе могут присутствовать или отсутствовать примечания любого типа.

Полезны редакционные примечания такого типа: «Проверить этот термин после поглощения компании» или «Этот термин упомянут в примечании к Понятию X» (см. 5.3) или «Проверить правописание с экспертом А». Такие примечания и некоторые атрибуты полезны не столько для пользователей, сколько для редакционной деятельности. Нужно сознательно подойти к решению вопроса о том, делать ли примечания видимыми для пользователя.

15.2.15 Ссылки примечаний

Связь между *Note* и *ThesaurusConcept* позволяет в примечании к одному понятию сделать ссылку на другое. Эта возможность особенно ценна для примечаний о значении термина.

15.2.16 Пользовательские атрибуты (Custom attributes)

Модель включает классы CustomConceptAttribute и CustomTermAttribute для пользовательских понятий и терминов, что позволяет записывать пользовательские данные о терминах и понятиях.

Они включаются как отдельные классы, чем обычные атрибуты, для того чтобы администратор системы ведения тезауруса мог устанавливать значение пользовательских атрибутов. Эти классы имеют атрибут customAttributeType, который позволяет администратору устанавливать, какой тип атрибута используется. Сами значения при этом обычно берутся из контрольного списка.

15.2.17 Метки рядов и узлов

Хотя все тезаурусы имеют ряды соподчиненных терминов, не всегда требуется моделировать их напрямую, и поэтому соответствующий класс факультативен. Классы ThesaurusArray и NodeLabel в модели необходимы в особенности для поддержки генерации представлений тезауруса, включающих метки узлов и/или неалфавитные последовательности соподчиненных терминов. (См. примеры на рисунках 4, 6, 7, 8, а также 12.4.3).

Атрибут *ordered* обеспечивает неалфавитный порядок, когда это необходимо.

В каждом языке тезауруса у каждого ряда дескрипторов должно быть не более одной метки узла.

Некоторые системы организуют выдачу так, как если бы метки узлов были понятиями с отношениями к вышестоящим и нижестоящим понятиям. Это делается только с целью визуализации, а не для нормального функционирования иерархических связей. Только понятия могут иметь настоящие связи к вышестоящим и нижестоящим понятиям.

15.2.18 Понятийные группы

Многие тезаурусы группируют понятия, используя классификационную структуру, которая существует независимо от параллельной иерархии понятий, основанных на отношениях выше-ниже. Порождаемые такой классификацией группы, как правило, базируются на научных дисциплинах, областях знаний или деловой активности. Их иногда называют «предметными категориями», «тематиками», «доменами», «группами» или «микротезаурусами». Настоящая модель все это предусматривает, предоставляя классы *ConceptGroup, ConceptGroupLabel*, а конкретный тип группировки может быть указан атрибутом *conceptGroupType*. Обычно между *ConceptGroup* и входящими в него понятиями тезаурусные отношения выше-ниже не устанавливаются.

В *ConceptGroup* могут быть собраны понятия из различных фасетов и иерархий тезауруса, а нотация классификации по группам может быть совершенно отлична от нотации понятий как таковых. Группы могут иметь подгруппы любого уровня. Каждой группе в каждом языке должна быть присвоена одна словесная метка.

Если требуется исключить определенные аскрипторы из данного множества понятий, может оказаться необходимым пометить их, что можно сделать с помощью пользовательских атрибутов (см. 15.2.16).

15.2.19 Фасеты

Фасеты предусмотрены моделью, но представлены в ней не напрямую, поскольку различные тезаурусы могут производить фасетный анализ по-разному. Если имена фасетов включены в качестве терминов верхнего уровня, их следует рассматривать как нормальные дескрипторы понятий верхнего уровня (см. 15.2.17). В противном случае, они могут представлять собой имена понятийных групп (15.2.18) или метки узлов (15.2.17).

15.2.20 Роли отношений

Классы *Equivalence, HierarchicalRelationship, AssociativeRelationship* имеют атрибут *role*. Он может быть использован вместе с отношениями С/См, В/Н и А/А. Его можно также использовать вместе с факультативными отношениями, описанными в п. 10.2.1, а также с пользовательскими отношениями, предложенными в п. 10.4. Каждое появление отношения выражает направление только в одну сторону, от экземпляра класса «источника» к экземпляру «целевого» класса. Например, «Понятие А имеет вышестоящим понятие В», где роль — «выше». Если дополняющее отношение «Понятие В имеет ни-

жестоящим понятие А» должно быть выражено, то оно указывается как отдельное появление того же отношения с ролью «ниже».

Для пользовательских типов отношений текст в атрибуте role должен состоять из (a) имени родительского типа отношений, следующего за ним (b) символа дроби «/» и (c) имени пользовательского типа отношений. При необходимости пользовательские типы отношений могут быть подразделены таким же способом.

Пример — Текст в атрибуте** role **класса** HierarchicalRelationship **может быть одним из указанных ниже, где NTX обозначает последующие подразделения NTI:
NT
NT/NTP
NT/NTI
NT/NTG
NT/NTI/NTX
BT
BT/BTP
BT/BTI
BT/BTG
BT/BTI/BTX
и т. д.

15.2.21 Расщепленные непредпочтительные термины (аскрипторы)

Классы *SplitNonPreferedTerm* и *CompoundEquivalence* обеспечивают представление сложных понятий посредством комбинации терминов, как то описано в 8.5.

15.2.22 Роль меток отношений

Рисунок 15 показывает отношение между *SimpleNonPreferedTerm* и *PreferedTerm* с метками USE и UF, которые описаны в разделах 3 и 8 в контексте соглашений человекочитаемого вывода. В контексте обмена данными это отношение может показаться излишним, поскольку все дескрипторы и аскрипторы данного понятия связаны с этим понятием. Однако модель сохраняет некоторую избыточность для облегчения обмена данными простых тезаурусов и для того, чтобы облегчить обмен данными обновлений и других специализированных подмножеств тезауруса.

Особенность представления отношения USE...+ состоит в том, что при человекочитаемом представлении между USE и + всегда имеется пробел для заполнения его термином (например, «coal mining USE coal + mining»). Но если тэг будет использоваться компьютерами за кулисами сцены, его лучше сжать в следующем виде: «coal mining USE+ coal; coal mining USE+ mining». По этой причине на диаграмме пробел отсутствует.

15.3 Табличное представление

Элементы модели, показанные на рисунке 15, перечислены в таблицах 4 — 9 по пяти классам — Thesaurus, ThesaurusArray, ThesaurusConcept, ThesaurusTerm и Note. Другие подчиненные классы показаны внутри этих таблиц. Краткий текст в колонке описания дополнен более распространенными примечаниями в 15.2, а также определениями или терминами, приведенными в разделе 2.

Колонка «Тип» показывает тип данных, используя где уместно типы «String» и «Boolean» по UML, «Date» и «Language» по XML и «Class» в случае начального или ассоциированного класса модели.

Колонка «Статус» показывает, является ли элемент обязательным и повторяемым согласно таблице 3.

Таблица 3 — Определение «Статуса» в таблицах 4 — 9

Статус	Значение
1	Обязательный, не повторяемый (только одно вхождение)
1..*	Обязательный, повторяемый (одно или более вхождений)
0..1	Факультативный, не повторяемый (нуль или одно вхождение)
0..*	Факультативный, повторяемый (нуль или более вхождений)

Таблица 4 — Атрибуты и ассоциации класса *Thesaurus*

Элемент	Тип	Статус	Описание
Thesaurus	Class	1	Тезаурус в целом
Атрибуты класса *Thesaurus*			
identifier	String	1..*	Идентификатор тезауруса в целом
contrbutor	String	0..*	Персона или организация, внесшая вклад в тезаурус
coverage	String	0..*	Пространственное или временно́е покрытие тезауруса
creator	String	0..*	Персона или организация, несущая основную ответственность за создание ресурса
date	String	0..*	Какая-либо дата, связанная с тезаурусом
created	String	0..1	Дата создания тезауруса
modified	String	0..*	Дата изменения тезауруса
description	String	0..*	Объяснение к тезаурусу
format	String	0..*	Формат файлов или физический носитель тезауруса
language	Language		Коды языков, поддерживаемых тезаурусом
publisher	String	0..*	Лицо, ответственное за публикацию (издание)
relation	String	0..*	Связанная публикация (издание)
rights	String	0..*	Копирайт или другая информация о правах
source	String	0..*	Ресурс, из данных которого произведен тезаурус
subject	String	0..*	Термины индексирования, определяющие тематику тезауруса
title	String	0..*	Наименование тезауруса
type	String	0..*	Вид словаря, например «thesaurus»
Ассоциированные классы класса *Thesaurus*			
ThesaurusArray	Class	0..*	Ряд родственных понятий, образующий часть тезауруса
ThesaurusConcept	Class	1..*	Понятие, входящее в тезаурус
ConceptGroup	Class	0..*	Группа понятий, выбранных по определенному критерию, например, по значимости в некоторой предметной области
VersionHistory	Class	0..*	Может подтверждать действенность версии, а также связь с другими версиями тезауруса

Таблица 5 — Атрибуты и ассоциации класса *ThesaurusArray*

Элемент	Тип	Статус	Описание
ThesaurusArray	Class	0..*	Ряд родственных понятий, входящих в тезаурус
Атрибуты класса *ThesaurusArray*			
identifier	String	1	Уникальный идентификатор ряда
ordered	Boolean	1	Метка истина/ложь, указывающая на необходимость сохранять порядок понятий в ряду. По умолчанию — «ложь»
notation	String	0..1	Символическое обозначение ряда, которое можно использовать для сортировки и визуализации

Окончание таблицы 5

Элемент	Тип	Статус	Описание
Ассоциированные классы класса *ThesaurusArray*			
hasSuperOrdinate ряд	Class	0..1	Ряд более высокого уровня, в который входит данный ряд
hasMember ряд	Class	0..*	Ряд более низкого уровня, входящий в данный ряд. *Ряд должен иметь по крайней мере один член, но он может быть либо рядом, либо понятием*
hasSuperOrdinate понятие	Class	0..1	Понятие более высокого уровня, которому подчинен данный ряд
hasMember понятие	Class	0..*	Понятие, являющееся членом данного ряда. *Ряд должен иметь по крайней мере один член, но он может быть либо рядом, либо понятием*
NodeLable	Class	0..*	Метка, показывающая основание, на котором выбраны и упорядочены понятия данного ряда
- Атрибуты класса *NodeLable*			
-lexicalValue	String	1	Словесное выражение метки узла
-created	Date	0..1	Дата создания метки узла
-modified	Date	0..1	Дата изменения метки узла
-lang	Language	0..1	Код, обозначающий язык метки узла. Используется в тех случаях, когда тезаурус поддерживает несколько языков

Таблица 6 — Атрибуты и ассоциации класса *ThesaurusConcept*

Элемент	Тип	Статус	Описание
ThesaurusConcept	Class	1..*	Понятие, входящее в тезаурус
Атрибуты класса *ThesaurusConcept*			
identifier	String	1	Уникальный идентификатор понятия
created	Date	0..1	Дата создания понятия
modified	Date	0..1	Дата изменения понятия
status	String	0..1	Указание, является ли понятие кандидатом, замененным и т. п.
notation	String	0..*	Символическое обозначение понятия, которое можно использовать для упорядочения и визуализации
topConcept	Boolean	0..1	Метка истина/ложь, указывающая на то, что понятие является понятием верхнего уровня, т. е. не имеет вышестоящих понятий
Ассоциированные классы класса *ThesaurusConcept*			
PreferedTerm	Class	1..*	Термин, служащий меткой данного понятия. В каждом языке у понятия должен быть только один дескриптор (предпочтительный термин)
SimpleNonPreferedTerm	Class	0..*	Альтернативный термин, по которому может быть найдено данное понятие
ScopeNote	Class	0..*	Примечание, определяющее или проясняющее значение термина в данном тезаурусе

Продолжение таблицы 6

Элемент	Тип	Статус	Описание
HistoryNote	Class	0..*	Примечание, фиксирующее изменения понятия в данном тезаурусе
CustomNote	Class	0..*	Примечание к понятию какого-либо другого типа
- Атрибуты класса *CustomNote*			
- *noteType*	String	0..1	Определение конкретного вида примечания
Note	Class	0..*	Примечание любого типа, которое ссылается на данное понятие, но более непросредственно прилагается к другому понятию
isMemberOfArray	Class	0..*	Ряд, в который входит данное понятие
hasSubordinateArray	Class	0..*	Ряд, подчиненный данному понятию
CustomConceptAtribute	Class	0..*	Дополнительный атрибут понятия
- Атрибуты класса *CustomConceptAtribute*			
- *lexicalValue*	String	1	Словесное обозначение атрибута
- *customAtributeType*	String	1	Определение конкретного вида атрибута
- *lang*	Language	0..1	Язык строки *lexicalValue*
Ассоциативный класс *TopLevelRelationship*			
- Классы, связанные посредством *TopLevelRelationship*			
- *hasTopConcept*	Class	0..*	Понятие наивысшего уровня, включающее данное понятие
- *isNopConceptOf*	Class	0..*	Понятие, на вершине иерархии которого находится данное понятие
Ассоциативный класс *HierarchicalRelationship*			
- Атрибуты класса *HierarchicalRelationship*			
- *role*	String	1	Определение вида иерархического отношения, которое связывает два понятия, например, BT, BTI, NT, NTP и др.
- Классы, связанные посредством *HierarchicalRelationship*			
- *hasHierRelConcept*	Class	0..*	Понятие, связанное с исходным понятием отношением, определенным в «role». Например, если «role» есть BT, то это — связь к вышестоящему понятию от данного понятия
- *isHierRelConcept*	Class	0..*	Исходное понятие, с которым данное понятие связано отношением, указанным в «role». Например, если роль есть BT, то это — связь к понятию, для которого данное понятие является вышестоящим
Ассоциативный класс *AssociativRelationship*			
- Атрибуты класса *AssociativRelationship*			
- *role*	String	0..1	Определение вида ассоциативного отношения, которое связывает два понятия. Это обычно RT, но отношение может быть несимметричным, например CAUSE, как уже обсуждалось в 10.4

Окончание таблицы 6

Элемент	Тип	Статус	Описание
- Классы, связанные посредством *AssociativRelationship*			
- *hasRelatedConcept*	Class	0..*	Понятие, связанное с данным понятием отношением, указанным в «role». Например, если отношение есть CAUSE, то это — связь с понятием, которое является причиной данного понятия
- *isRelatedConcept*	Class	0..*	Исходное понятие, с которым связано данное понятие отношением, указанным в «role». Например, если роль есть CAUSE, то это — связь с понятием, для которого данное понятие является причиной

Таблица 7 — Атрибуты и ассоциации класса *ConceptGroup*

Элемент	Тип	Статус	Описание
ConceptGroup	Class	0..*	Группа понятий, выбранная по некоторому критерию, такому как актуальность в определенной предметной области
Атрибуты класса *ConceptGroup*			
identifier	String	1	Уникальный идентификатор группы
conceptGroupType	String	1	Метка, указывающая тип группы, например, «микротезаурус», «тематика» или «предметная категория»
notation	String	0..1	Символическое обозначение группы, которое можно использовать для сортировки и визуализации
Ассоциированные классы класса *ConceptGroup*			
ThsaurusConcept	Class	0..*	Понятие, входящее в группу
hasSuperGroup	Class	0..*	Группа более высокого уровня, членом которой является данная группа
hasSubGroup	Class	0..*	Группа нижнего уровня, являющаяся членом данной группы
ConceptGroupLabel	Class	1..*	Метка, представляющая собой словесное описание группы. Группа в каждом языке должна иметь одну такую метку
- Атрибуты класса *ConceptGroupLabel*			
- *lexicalValue*	String	1	Словесное выражение метки
- *created*	Date	0..1	Дата создания метки
- *modified*	Date	0..1	Дата изменения метки
- *lang*	Language	0..1	Код, обозначающий язык метки. Используется в тех случаях, когда тезаурус поддерживает несколько языков

Таблица 8 — Атрибуты и ассоциации класса *ThesaurusTerm*

Элемент	Тип	Статус	Описание
ThesaurusTerm	Class	1	Термин тезауруса, по которому может быть найдено понятие
Атрибуты класса *ThesaurusTerm*			
lexicalValue	String	1	Словесное представление термина
identifier	String	1	Уникальный идентификатор термина
created	Date	0..1	Дата создания термина

Продолжение таблицы 8

Элемент	Тип	Статус	Описание
modified	Date	0..1	Дата изменения термина
source	String	0..1	Лица или документы, откуда был получен тезаурус
status	String	0..1	Указание, является ли термин кандидатом, замененным и т. п.
lang	Language	0..1	Код языка термина. Должен указываться, если тезаурус поддерживает несколько языков
Ассоциативные классы класса *ThesaurusTerm*			
HistoryNote	Class	0..*	Примечание, фиксирующее изменения термина в тезаурусе
EditorialNote	Class	0..*	Примечание для редакторов, используемое при редактировании тезауруса
Definition	Class	0..*	Примечание, дающее определение термина, не обязательно ограниченное значением понятия, обозначаемого этим термином в данном тезаурусе
- Атрибуты класса *Definition*			
- *source*	String	0..1	Лица или документы, откуда было получено определение
CostomTermAttribute	Class	0..*	Дополнительный атрибут термина
- Атрибуты класса *CostomTermAttribute*			
- *lexicalValue*	String	1	Словесное представление атрибута
- *CostomAttributType*	String	1	Указание на конкретный вид атрибута
- *lang*	Language	0..1	Язык строки *lexicalValue*
Специализации класса *ThesaurusTerm*			
PreferedTerm	Class		Термин, выбранный в качестве метки понятия
SimpleNonPreferedTerm	Class		Термин, который служит как альтернативная метка понятия, но не является дескриптором
- Атрибуты класса *SimpleNonPreferedTerm*			
- *hidden*	Boolean	0..1	Флажок да/нет, показывающий, можно ли термин исключать из некоторых видов выходных форм
SplitNonPreferedTerm	Class		Термин, который представляет комбинацию двух или более понятий тезауруса
- Ассоциации классов *PreferedTerm* и *SimpleNonPreferedTerm*			
- *Equivalence*	Association class		
- - Атрибуты класса *Equivalence*			
- - *role*	Class	0..1	Уточнение типа отношения эквивалентности. Обычно это USE, связывающий исходный *SimpleNonPreferedTerm* с *PreferedTerm*
- - Классы, связанные посредством *Equivalence*			
- - *PreferedTerm*	Class	1	Термин, связанный с данным исходным термином отношением, указанным в «role». Напр. если роль — USE, то это ссылка на дескриптор, который следует использовать вместо данного аскриптора

Окончание таблицы 8

Элемент	Тип	Статус	Описание
- - *SimpleNonPrefered-Term*	Class	0..*	Исходный термин отношения, с которым данный термин связан отношением, указанным в «role». Напр. если роль — USE, то это ссылка на аскриптор, вместо которого употребляется данный дескриптор
- - Ассоциации классов *PreferedTerm* и *SplitNonPreferedTerm*			
- - *CompoundEquivalence*	Association class		
- - Классы, связанные посредством *CompoundEquivalence*			
- - *SplitNonPreferedTerm*	Class	0..*	Аскриптор, именующий сложное понятие, представленное комбинацией дескрипторов. См. 8.5
- - *PreferedTerm*	Class	2..*	Один из двух или нескольких дескрипторов, которые совместно представляют данный *SplitNonPreferedTerm*[1)]

Таблица 9 — Атрибуты и ассоциации класса *Note*

Элемент	Тип	Статус	Описание
Note	Class	0..*	Текст, содержащий дополнительные сведения о термине или понятии
Атрибуты класса *Note*			
lexicalValue	String	1	Словесное представление текста примечания
created	Date	0..1	Дата создания примечания
modified	Date	0..1	Дата изменения примечания
lang	Language		Код, указывающий язык примечания. Используется в случае, если тезаурус поддерживает несколько языков
Ассоциированные классы класса *Note*			
TesaurusConcept	Class	0..*	Понятие, на которое ссылается примечание, но не то, к которому примечание напрямую привязано
Специализации отдельных типов класса *Note*			
ScopeNote	Class	0..*	Примечание, определяющее или разъясняющее понятие
HistoryNote	Class	0..*	Примечание, фиксирующее изменения понятия или термина
Definition	Class	0..*	Определение термина, не обязательно ограниченное понятиями, представляемыми терминами тезауруса
EditorialNote	Class	0..*	Примечания для использования при редактировании тезауруса
CustomeNote	Class	0..*	Примечание какого-либо другого типа к понятию
- Атрибуты класса *CustomeNote*			
- *noteType*	String	0..1	Указание конкретного типа примечания

16 Интеграция тезаурусов с приложениями

16.1 Введение

Первичная функция тезауруса — это помощь в нахождении информации. Тезаурус обычно используют на двух этапах процедуры нахождения: во время индексирования и во время поиска. Хотя существуют программные пакеты, которые управляют созданием, развитием и ведением тезауруса,

[1)] В оригинале ошибочно указано «CompoundNonPreferedTerm».

индексированию и поиском в одной интегрированной системе, чаще бывает принято рассматривать эти функции как отдельные приложения. Поэтому необходимо, чтобы тезаурусными данными можно было беспрепятственно обмениваться между этими приложениями, при этом следует избегать ошибок и неопределенностей при текущем обновлении данных.

С распространением электронных сетей и почти всеобщим использованием компьютеров в управлении информацией сценарий становится все более сложным, что дает возможность обмена данными и информацией в масштабе, который раньше нельзя было даже вообразить. Из одного компьютера, включенного в сеть, можно одновременно опрашивать различные коллекции книг, статей, фотографий, наборов данных и музейных экспонатов, размещенные на компьютерах, распределенных по всему миру. Чтобы обеспечить возможность использования тезауруса в подобном сценарии, от систем требуется, чтобы системы были совместимыми, т. е. разные системы или их компоненты были способны обмениваться информацией для внутрисистемного использования.

В настоящем разделе описываются:

a) свойства и функции, которые обеспечивают совместимость тезауруса с системами и сетями;

b) свойства и функции, которые необходимы программам индексирования и поиска, использующим тезаурус.

Здесь не описываются свойства и функции, необходимые таким приложениям, которые используют два или более связанных тезауруса. Такие требования будут включены в [6] и следующую версию стандарта.

16.2 Совместимость, необходимая для тезаурусов

Как описано в 14.8, программное обеспечение, используемое для ведения тезауруса, должно быть способно экспортировать и импортировать тезаурус целиком и по частям. Для этого должен использоваться стандартный формат тезауруса (см. раздел 17).

Для замены старой версии тезауруса новой должна быть предусмотрена возможность экспортировать все новые, измененные и исключенные статьи.

Когда внедряется новая версия тезауруса, система должна быть в состоянии включить все новые или измененные термины без потери данных, приписанных к сохранившимся терминам, если они еще действительны. Например, отсылки от терминов к ресурсам, которые заиндексированы этими терминами.

Для предоставления возможности использования тезауруса пользователями сети (например, Интернета) тезаурус должен быть опубликован в сети в стандартном формате (см. раздел 17).

Опубликованный тезаурус должен регулярно обновляться. Это может потребовать импорта и загрузки файла обновлений из приложения, в котором поддерживается эталонная версия.

Если тезаурус опубликован в сети, для получения всех или части данных тезауруса должен использоваться стандартный протокол (см. раздел 18). Например, программе поиска может потребоваться выбрать всю или часть информации, связанной с отдельным понятием или понятийной группой. С каждым понятием должен быть связан уникальный идентификатор, что облегчает распределенный поиск.

Для удобства пользователей и широкого распространения тезаурус должен быть зарегистрирован в каком-либо словарном регистре, который обслуживает целевую аудиторию пользователей.

16.3 Интеграция с программами индексирования и поиска

16.3.1 Если тезаурус предназначен для информационного поиска, он должен быть совместим с системами, в которых происходит индексирование и поиск. Для печатных тезаурусов интеграция достигается тем, что человек, просматривая страницы, выбирает термины и вводит их в информационно-поисковую систему с соответствующим синтаксисом. Для электронных тезаурусов эти функции должны поддерживаться дружественным интерфейсом.

Пока ведение тезауруса не объединено в одной системе с индексированием и поиском, первое требование состоит в том, чтобы экспортировать тезаурусные данные из системы ведения тезауруса в информационно-поисковую систему. Форматы и протоколы для этого описаны в разделах 17 и 18.

16.3.2 Все системы, которые вовлечены в обработку тезаурусных данных, должны поддерживать все символы, которые могут быть представлены в данных. Они обычно включают скобки квалификаторов терминов и иногда другие знаки препинания. В случае многоязычного тезауруса должны поддерживаться все символы письменностей и алфавитов всех его языков.

16.3.3 Для программ индексирования и поиска, за исключением тех, которые полностью автоматизированы и не предусматривают вмешательства пользователя, необходимо обеспечить возможности просмотра и поиска в тезаурусе. Это включает следующее:

a) Возможность поиска термина, содержащего какое-нибудь слово, словосочетание или буквосочетание.

b) Желательно, чтобы была предоставлена помощь поиска термина и понятия по обозначению или имеющимся идентификаторам.

c) Возможность выбирать и переключать язык интерфейса многоязычного тезауруса, а также ограничивать поиск и выдачу одним из языков.

d) Возможность показывать эквиваленты для данного термина, дескриптора (предпочтительного термина) или аскриптора (непредпочтительного термина), включая иноязычные эквиваленты, если тезаурус многоязычный.

e) Возможность переходить от записи для одного понятия к записи для вышестоящих, нижестоящих и ассоциативных понятий, что обычно достигается установлением гиперссылок. Эта возможность может исключить потребность в нотации для связи алфавитного указателя с классификационным или иерархическим.

f) Возможность увеличивать или сокращать глубину просмотра классификационного или иерархического указателя. Для получения начального обзора структуры должна быть предусмотрена возможность начинать с просмотра верхнего уровня, а затем выборочно увеличивать глубину просмотра.

g) Возможность переключать стили просмотра, например от классификационного к алфавитному отображению. Предпочтительно, чтобы была возможность одновременного просмотра обоих этих указателей.

16.3.4 Когда тезаурус интегрирован с поисковой коллекцией документов и имеется необходимость, чтобы пользователь консультировался с тезаурусом, на экране также должно отображаться число использований каждого дескриптора (предпочтительного термина). При просмотре тезауруса пользователь должен иметь возможность перейти прямо к фазе поиска. Должны быть предоставлены следующие возможности:

a) Возможность выбрать для поиска один или более терминов, отображенных на экране.

b) Возможность создать поисковое предписание посредством копирования терминов из тезауруса без необходимости повторного набора текста (например, посредством двойного щелчка или перетаскивания). Идеально будет, если будет предоставлена помощь по включению синтаксиса (например, по выбору булевских операторов).

c) Возможность выбрать для поиска иерархические поддеревья тезауруса, т. е. искать по данному термину и по всем нижестоящим к нему терминам, соединенным булевским ИЛИ, что иногда называется «взрывным» поиском.

d) Выбор аскриптора (непредпочтительного термина) должен вести к замене его в поисковом предписании соответствующим дескриптором (предпочтительным термином), при обеспечении пользователя сведениями об этом. В том случае, когда пользователь ввел в поисковое предписание аскриптор (непредпочтительный термин) без предварительного просмотра тезауруса, ему должно быть предложено конвертировать его в соответствующий дескриптор (предпочтительный термин). Во время процедур улучшения поисковых характеристик в каждом из указателей тезауруса число использований терминов должно соответствовать действительному положению для выбранной коллекции данных.

16.3.5 Тезаурус также можно использовать в поисковой машине без предъявления его пользователю для просмотра. Для этого типа использования могут быть предусмотрены возможности, указанные в 16.3.4, перечисления c) и d).

Примеры —

A) Если пользователь вводит термин, сходный с одним или несколькими терминами тезауруса, поисковая машина может ответить «Вы имели в виду "..."?», таким способом предлагая пользователю ближайшие подходящие дескрипторы тезауруса.

B) Если пользователь вводит термин, сходный с несколькими терминами тезауруса, он может в ответ найти результаты по всем этим терминам и рассортировать выдачу по этим терминам. (Например, если тезаурус имеет два термина «turkey (meat)» и «Turkey (country)», а пользователь ищет просто «Turkey», результаты поиска могут быть представлены двумя списками, соответствующими каждому из терминов тезауруса.)

C) Пользователю может быть предоставлена возможность «расширять» и «сужать» поиск без объяснения, как это делается.

D) При получении недостаточной выдачи поиск может быть автоматически расширен вышестоящими, нижестоящими и ассоциативными терминами.

Если поисковые термины автоматически конвертируются, замещаются, переводятся, заменяются вышестоящими или нижестоящими терминами, пользователь должен быть проинформирован, что было сделано.

16.3.6 Отношения эквивалентности в тезаурусе позволяют производить поиск по базам данных, которые не индексировались по данному тезаурусу. Например, если тезаурус включает термины «motor-bikes», «motor-cycles» и «motorcycles» в качестве дескрипторов или аскрипторов (предпочтительных или непредпочтительных терминов) для одного и того же понятия, то поиск по любому из этих терминов может быть расширен включением поиска по всем остальным синонимам. См. также обсуждение синонимических колец (synonym rings) в [6].

16.3.7 При интеграции в систему индексирования или каталогизации должны быть обеспечены следующие функции:

a) копирование одного или более терминов прямо в формуляр индексирования без необходимости вводить слова с клавиатуры;

b) проверка по тезаурусу терминов, введенных напрямую (при вводе напрямую аскриптора непредпочтительного термина), он должен быть заменен на дескриптор автоматически с извещением об этом пользователя. При вводе строки символов, совпадающей с началом одного или более терминов тезауруса, индексатору должны быть предложены для выбора все соответствующие дескрипторы (предпочтительные термины);

c) возможность ввода в тезаурус кандидатов в дескрипторы для рассмотрения их редактором тезауруса (см. 13.6).

16.3.8 Если система предназначена для автоматического индексирования, а не для работы специалиста, то должны быть предусмотрены средства связи каждого дескриптора (предпочтительного термина) с алгоритмом или правилами, которые управляют его использованием.

17 Обменные форматы

Взаимодействие различных компьютерных приложений требует использования общего формата. Хорошо известны следующие четыре формата:

— **MARC** (Mashine-Readable Cataloging) [22] — Этот набор стандартов, основанный на ИСО 2709, широко используется для обмена библиографическими данными между библиотеками. Хотя основное применение этих стандартов относится к обмену каталожными записями, в набор входят также стандарты по обмену «авторитетными записями», т. е. записями, взятыми из тезаурусов и других типов контрольных словарей. Особое значение имеют:

1) MARC 21 Формат для авторитетных данных (имеется в полной и сокращенной версии) [23];

2) MARC 21 Формат для классификационных данных (имеется в полной и сокращенной версии) [24];

3) MADS (Metadata Authority Description Schema) — совместимый с MARC 21 формат описания на XML данных того типа, которые содержатся в записях авторитетных данных.

Форматы MARC, выраженные на языке XML, соответствуют [25].

- **SKOS** — Формат SKOS (Simple Knowledge Organization Sistems) [26, 27] установлен рекомендацией W3C применительно к работам по семантическому вэбу и использует кодировку XML и RDF (Resource Description Framework). Он разработан как для тезаурусов, так и для классификационных систем, таксономий и других типов словарей.

- **Zthes** — Zthes [28] первоначально был разработан как прикладной профиль Z39.50 для обмена тезаурусными данными, но впоследствии был развит для использования с SRU (Search and Retrieval via URL; см. раздел 18).

- **DD 8723-5** — Этот проект британского стандарта предлагает схему XML, которую можно использовать для обмена тезаурусом, удовлетворяющим британскому стандарту BS 8723 [29]. Схема доступна на gost.e-ecolog.ru

Дополнительно к этим четырем форматам для удовлетворения потребностей отдельных сфер применяют и другие форматы.

В частности, находит применение [30] — Система разметки терминологических данных TMF (Terminological Markup Framework) для компьютерных работ в терминологии. Эта схема разработана для обеспечения единых принципов представления данных в терминологических коллекциях. Она включа-

ет метамодель и методы описания специализированных терминологических языков разметки (TML), выраженных форматом XML. Эта разработка поддерживает использование и развитие компьютерных приложений к терминологическим данным и обмен ими между различными приложениями. Категории данных и методы описания структур данных определены в [31] и других родственных международных стандартах ИСО.

Поскольку ни один из существующих форматов не способен передавать все особенности тезаурусов, использующих модель раздела 15, рекомендуется формат XML.

Похоже на то, что в течение некоторого времени будут сосуществовать несколько различных форматов. Один из путей минимизации неудобств множественности форматов состоит в том, чтобы конвертировать данные подходящим преобразователем, даже если некоторая информация будет теряться, если тезаурус преобразуется в формат, не поддерживающий все особенности исходного формата. В настоящее время разрабатывается несколько таких преобразователей. Сайт Schemas (gost.e-ecolog.ru niso.org/schemas/iso25964) предоставляет место, на котором могут быть найдены будущие разработки.

18 Протоколы

18.1 Общие сведения

Протокол является важным инструментом коммуникационного процесса, позволяющим использовать тезаурусы в компьютерных сетях. Используется несколько протоколов, обеспечивающих быстрое получение сведений из тезаурусов с целью индексирования или поиска, как то описано в п. 16.3. В этой быстро развивающейся области постоянно появляются новые разработки, поэтому создать полный перечень протоколов невозможно. Поскольку каждый протокол имеет силу в своих обстоятельствах и в окружении специфических приложений, то и те, которые перечислены в п. 18.4 и п. 18.5, тоже могут рассматриваться только для применения в определенных приложениях.

18.2 Цели и случаи использования

Важно отличать представление тезауруса в целом с его метаданными от поиска отдельных понятий и терминов (или групп понятий и терминов). Протоколы могут быть использованы для представления и применения тезаурусов, содержащихся в них понятий, терминов и отношений, для описания смысла терминов и облегчения семантической совместимости. Это делается для широкого спектра задач поиска, просмотра, перевода, сопоставления, семантического обоснования, предметного индексирования, классификации, сбора данных, оповещения [32].

18.3 Окружение и архитектура приложений

Некоторые протоколы разработаны специально для тезаурусов (см. 18.4), и в то же время используются (или адаптируются) другие протоколы общего назначения (см. 18.5). Специализированные протоколы позволяют более интенсивно использовать структуру тезауруса и отношений, а протоколы общего назначения могут облегчить интеграцию с другими функциями и возможностями систем. Также имеется возможность дополнить протокол общего назначения тезаурусным интерфейсом. В некоторых случаях ограниченный набор функциональных команд, предусмотренный определенным тезаурусным интерфейсом, предоставляет удобство тем, что скрывает детали архитектуры и реализации функций. С другой стороны, иногда большая гибкость протоколов общего назначения может быть с пользой применена квалифицированным разработчиком.

Для отдельных приложений может оказаться необходимым принять в расчет характер платформы, на которой реализуется тезаурус, например SOAP [33] или более «легкий» HTTP протокол REST [34] (и другие того же типа) XML-RPC [35] или JSON-RPC [36]. Некоторые тезаурусы недавно были опубликованы как Liked Data [37] — механизм связи, основанный на использовании RDF и URI, что можно рассматривать как вариант REST.

18.4 Специализированные тезаурусные протоколы

Заслуживают упоминания три протокола:

— **SWAD-E SKOS API** — Это протокол программирования приложений (Web Service API), предназначенный для обеспечения доступа к тезаурусам и другим простым системам организации знаний (SKOS) через Интернет. Он определяет ядерный набор операций для программируемого доступа и

опроса тезауруса. Рассчитанный на сетевые запросы, сам API сохраняет независимость от деталей конкретной реализации. Этот протокол может быть адаптирован и к форматам, отличным от SKOS. В то время как SKOS API — результат работ по европейскому проекту Semantic Web Advanced Development (SWAD — Перспективное развитие семантического вэба), сам SKOS поддерживается рабочей группой Semantic Web Deploiment (SWDWG — Рабочая группа развертывания семантического вэба).

— **Другие API** — В настоящее время используют множество различных адаптаций SKOS API на различных платформах. Существуют также похожие, но независимые тезаурусные протоколы, основанные как на SOAP, так и на REST. Их списки, не претендующие на полноту, ведутся по ссылкам [38] и [39].

— **ADL Thesaurus Protocol** — Протокол библиотеки Alexandria Digital Library (ADL) рассчитан на доступ к одноязычному тезаурусу и адаптирован к использованию через Интернет. Он поддерживает оперативный опрос тезауруса и навигацию по нему. Он предлагает возможность «легкого» HTTP доступа к тезаурусу. Модель тезауруса, используемая в протоколе, близко следует стандарту ANSI/NISO Z39.19 [40], а его определение дано по схеме XML.

18.5 Протоколы баз данных общего назначения, используемые с тезаурусами

Для некоторых приложений могут быть полезны следующие протоколы общего назначения:

— **Search Web Services (OASIS)** — Эти протоколы основаны на ADP (Abstract Protocol Definition) [41] — эталонная модель, которая способна описывать эти и другие протоколы. Наиболее заметными являются [42] и SRU (Search and Retrieval via URL) [43] с поисковым языком CQL (Contextual Query Language) [44]. Рассчитанный на использование с неструктурированными документами OpenSearch проще, а SRU обеспечивает больше возможностей по управлению запросами и выдачей. SRU сфокусирован на XML запросах в Интернете. SRU/CQL в настоящее время пересматривается Техническим комитетом OASIS [45]. Более ранний протокол, известный как SRW, теперь рассматривается как вариант SRU ("SRU viaHTTP SOAP"); SRU основан на протоколе REST, а SRW — на протоколе SOAP.

— **SPARQL** (Simple Protocol and RDF Query Language — Простой протокол и язык запросов RDF) [1] — Установлен официальной рекомендацией Консорциума WWW (W3C) 2008 года. Он разработан и продвигается консорциумом как ключевая технология семантического вэба. Язык запросов RDF используется для выражения запросов к различным источникам данных, как в тех случаях, когда данные записаны исконно в RDF, так и в случаях, когда они рассматриваются через транслятор. SPARQL обладает возможностью поиска требуемых и возможных графических образов, а также их соединений и разъединений. Его можно рассматривать как язык запросов общего назначения (наподобие SQL для реляционных баз данных), и он может быть использован для опроса тезаурусов, представленных в RDF, совместно с другими наборами данных в том же формате.

— **Z39.50** — Этот протокол более формально обозначается как "ANSI/NISO Z39.50 Information retrieval (Z39.50) Application service definition and protocol specification" [46]. SRU/SRW можно рассматривать как варианты ранних версий протокола Z39.50.

Библиография

[1] World Wide Web Consortium. SPARQL Protocol for RDF. W3C Recommendation, 15 January 2008. Available at: gost.e-ecolog.ru

[2] World Wide Web Consortium. XML Schema Part 0: Primer Second Edition. W3C Recommendation, 28 October 2004. Available at gost.e-ecolog.ru

[3] World Wide Web Consortium. XML Schema Part 1: Structures Second Edition. W3C Recommendation, 28 October 2004. Available at gost.e-ecolog.ru

[4] World Wide Web Consortium. XML Schema Part 2: Datatypes Second Edition. W3C Recommendation, 28 October 2004. Available at gost.e-ecolog.ru

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

[6] ИСО 25964—2:2013 «Информация и документация. Тезаурусы и взаимосвязь с другими словарями. Часть 2. Взаимосвязь с другими словарями»: ISO 25964-2:2013 Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies

[7] ИСО 25964.1 (ISO 25964-1:2011 «Information and documentation — Thesauri and interoperability with other vocabularies — Part 1. Thesauri for information retrieval», MOD

[8] ГОСТ 7.79 — 2000 «Система стандартов по информации, библиотечному и издательскому делу. Правила транслитерации кирилловского письма латинским алфавитом»

[9] ГОСТ Р 7.0.34 — 2014 «Система стандартов по информации, библиотечному и издательскому делу. Правила упрощенной транслитерации русского письма латинским алфавитом»

[10] Joint Steering Committee for Revision of AACR. Anglo-American cataloguing rules, 2nd ed., 2002 revision with 2004 update. London: Facet Publishing for Chartered Institute of Library and Information Professionals, 2004. ISBN 0-85604-469-6

[11] ГОСТ 7.24 — 2007 «Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый многоязычный. Состав, структура и основные требования к построению»

[12] ISO 8601, Data elements and interchange formats — Information interchange — Representation of dates and times

[13] ISO/IEC 10646, Information technology — Universal Coded Character Set (UCS)

[14] Object Management Group. Unified Modeling Language™: UML® resource page. Needham, MA: Object Management Group, 2007. [Available at gost.e-ecolog.ru

[15] Object Management Group. Unified Modeling Language (UML), version 2.1.2. Needham, MA: Object Management Group, 2007. Available at gost.e-ecolog.ru

[16] Object Management Group. Unified Modeling Language Specification, version 1.4.2. Needham, MA: Object Management Group, January 2005. Available at gost.e-ecolog.ru [Also available from ISO as ISO/IEC 19501:2005]

[17] Dublin Core Metadata Initiative. Dublin core metadata element set, version 1.1. DCMI recommendation, 18 December 2006. Latest version available at gost.e-ecolog.ru [Full set of DCMI standards available at gost.e-ecolog.ru]

[18] ISO 639 — 1:2002 Codes for the representation of names of languages — Part 1: Alpha-2 code

[19] ISO 639 — 2:1998 Codes for the representation of names of languages — Part 2: Alpha-3 code

[20] Phillips, A. and M. Davis, eds. Tags for identifying languages, RFC 4646. The Internet Society, September 2006. Available at gost.e-ecolog.ru

[21] Internet Assigned Numbers Authority. Language Subtag Registry. Available at: gost.e-ecolog.ru language-subtag-registry

[22] US Library of Congress Network Development and MARC Standards Office. MARC Standards. gost.e-ecolog.ru marc/authority/

[23] US Library of Congress Network Development and MARC Standards Office. MARC 21 format for authority data. Washington, DC: Library of Congress. Concise version available online at gost.e-ecolog.ru

[24] US Library of Congress Network Development and MARC Standards Office. MARC 21 format for classification data. Washington, DC: Library of Congress. Concise version available online at gost.e-ecolog.ru

[25] ИСО 25577 Информация и документация. Электронный формат MarcXchange: Information and documentation — MarcXchange

[26] World Wide Web Consortium. SKOS Simple Knowledge Organization System Reference. W3C Recommendation, 18 August 2009. Latest version available at gost.e-ecolog.ru

[27] World Wide Web Consortium. SKOS Simple Knowledge Organization System Primer. W3C Working Group Note, 18 August 2009. Latest version available at gost.e-ecolog.ru

[28] Zthes Working Group. The Zthes specifications for thesaurus representation, access and navigation. 17 February 2006. Available at gost.e-ecolog.ru

[29] BS 8723, Structured vocabularies for information retrieval — Guide

[30] ИСО 16642—2003 Применение компьютера в терминологических целях. Структура терминологической разметки: Computer applications in terminology — Terminological markup framework

[31] ИСО 12620—2012 Терминология, другие языковые ресурсы и ресурсы содержания. Сецификация категорий данных и ведение реестра категорий данных для языковых ресурсов: ISO 12620:2009 Terminology and other language and content resources — Specification of data categories and management of a Data Category Registry for language resouces

[32] Tudhope, Douglas, Traugott Koch, and Rachel Heery. Terminology Services and Technology: JISC state of the art review. Bath, UK: UKOLN, 15 September 2006. Available at: gost.e-ecolog.ru capital/terminology_services_and_technology_review_sep_06.pdf

[33] World Wide Web Consortium. SOAP Version 1.2. W3C Recommendation, 27 April 2007. Available at: gost.e-ecolog.ru w3.org/TR/soap12

[34] Fielding, R. T. “Representational State Transfer (REST).” In: Architectural Styles and the Design of Network-based Software Architectures, Chapter 5. University of California, Irvine Ph.D. Dissertation, 2000. Available at: gost.e-ecolog.ru ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm

[35] XML-RPC [website]. UserLand Software, Inc. Available at: gost.e-ecolog.ru

[36] JSON-RPC Specification. JSON-RPC.ORG, 2005. Available at: gost.e-ecolog.ru

[37] Linked Data [website]. Available at: gost.e-ecolog.ru

[38] KOS-based web services. Pontypridd, UK: University of Glamorgan. Available at: gost.e-ecolog.ru ac.uk/kos/terminology_services/links/

[39] STITCH (Semantic Interoperability to Access Cultural Heritage). Repository — Related Work [webpage]. Available at: gost.e-ecolog.ru

[40] ANSI/NISO Z39.19:2005, Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Available at: gost.e-ecolog.ru

[41] Denenberg, Ray. “Search Web Services — The OASIS SWS Technical Committee Work: The Abstract Protocol Definition, OpenSearch Binding, and SRU/CQL 2.0.” D-Lib Magazine, 15 (1/2), January/February 2009. Available at: gost.e-ecolog.ru

[42] Clinton, DeWitt. OpenSearch 1.1, draft 3. Palo Alto, CA: A9.com, Inc. Available at: gost.e-ecolog.ru

[43] SRU: Search/Retrieve via URL, version 1.2. Washington, DC: Library of Congress. Available at gost.e-ecolog.ru standards/sru/

[44] CQL: Contextual Query Language. Washington, DC: Library of Congress. Available at: gost.e-ecolog.ru sru/specs/cql.html

[45] OASIS Search Web Services Technical Committee [webpage]. Available at: gost.e-ecolog.ru tc_home.php?wg_abbrev=search-ws

УДК 025.43:006.72(083.74) ОКС 01.140.20 Т62

Ключевые слова: тезаурус, информационный поиск, иерархические отношения, ассоциативные отношения, семантическая эквивалентность терминов, дескрипторы, аскрипторы, логические отношения понятий, форматы представления тезаурусов

Редактор *О.А. Антошкова*
Корректор *М.В. Бучная*
Компьютерная верстка *Л.А. Круговой*

Подписано в печать 08.02.2016. Формат 60 ×84 1/8.
Усл. печ. л. 11,16. Тираж 50 экз. Зак. 307.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

ФГУП «СТАНДАРТИНФОРМ», 123995 Москва, Гранатный пер., 4.
gost.e-ecolog.ru gost.e-ecolog.ru

gost.e-ecolog.ru