ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 7.0.66— 2010 (ИСО 5963:1985)

Система стандартов по информации, библиотечному и издательскому делу

ИНДЕКСИРОВАНИЕ ДОКУМЕНТОВ

Общие требования к координатному индексированию

(ISO 5963:1985,
Documentation — Methods for examining documents, determining their subjects and selecting indexing terms,
MOD)

Издание официальное

Москва
Стандартинформ
2019

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН Всероссийским институтом научной и технической информации Российской академии наук (ВИНИТИ РАН) и ФГУП «ВИМИ» на основе собственного перевода на русский язык англоязычной версии 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 9 ноября 2010 г. № 344-ст

4 Настоящий стандарт включает в себя модифицированные основные нормативные положения международного стандарта ИСО 5963:1985 «Документация. Методы анализа документов, определения их темы и подбора индексирующих терминов» (ISO 5963:1985 «Documentation — Methods for examining documents, determining their subjects and selecting indexing terms», MOD)

Наименование настоящего стандарта изменено относительно наименования указанного международного стандарта для приведения в соответствие с ГОСТ Р 1.5—2012 (пункт 3.5)

5 ВВЕДЕН ВПЕРВЫЕ

6 ПЕРЕИЗДАНИЕ. Январь 2019 г.

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». 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

© ISO, 1985 — Все права сохраняются
© Стандартинформ, оформление, 2011, 2019

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

Содержание

1 Область применения . . . 1
2 Нормативные ссылки . . . 1
3 Термины и определения . . . 2
4 Общие положения . . . 2
5 Анализ документа . . . 2
6 Выбор понятий, характеризующих содержание документа . . . 3
7 Выбор терминов индексирования . . . 3
8 Формирование поискового образа документа . . . 4
9 Фактографическое индексирование документа . . . 6
10 Автоматизированное индексирование . . . 7
Приложение А (справочное) Схема индексирования по информационно-поисковому тезаурусу . . . 9

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Система стандартов по информации, библиотечному и издательскому делу

ИНДЕКСИРОВАНИЕ ДОКУМЕНТОВ

Общие требования к координатному индексированию

System of standards on information, librarianship and publishing. Indexing of documents.
General requirements for coordinate indexing

Дата введения — 2011—07—01

1 Область применения

Настоящий стандарт устанавливает общие требования к координатному индексированию документов, включая правила формирования поискового образа документа. Специфические требования к систематизации и предметизации документов — по *ГОСТ 7.59*. Форма представления поискового образа документа — по *ГОСТ Р 7.0.52*.

Настоящий стандарт распространяется на информационно-поисковые системы, в которых содержание документов представлено в сжатой форме лексическими единицами информационно-поискового языка. Настоящий стандарт не распространяется на формирование фактографических записей в фактографических базах данных.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р 7.0.52 Система стандартов по информации, библиотечному и издательскому делу. Формат для обмена библиографическими данными. Поисковый образ документа

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

ГОСТ 7.59 Система стандартов по информации, библиотечному и издательскому делу. Индексирование документов. Общие требования к систематизации и предметизации

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

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

Издание официальное

3 Термины и определения

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

3.1 **автоматизированное индексирование:** Индексирование, технология которого предусматривает использование формальных процедур, осуществляемых с помощью вычислительной техники, и может включать в себя применение интеллектуальных процедур при принятии основных решений о составе поискового образа.

3.2 **автоматическое индексирование:** Составление поискового образа с использованием только формальных процедур обработки текста документа или запроса, осуществляемых средствами вычислительной техники.

3.3 **ключевое слово:** Слово или словосочетание (термин) в тексте документа или запроса, несущее в нем существенную информационную нагрузку, хотя бы по одной из тем, рассматриваемых в документе.

3.4 **контролируемое индексирование (дескрипторное индексирование):** Индексирование, при котором ключевые слова текста заменяются дескрипторами, указанными в соответствующем дескрипторном словаре.

3.5 **координатное индексирование:** Индексирование, цель которого состоит во всестороннем отражении содержания документа или запроса путем включения в поисковый образ всех необходимых для этого терминов индексирования.

3.6 **свободное индексирование:** Индексирование, технология которого не предусматривает замену ключевых слов текста в соответствии с рекомендациями специального словаря индексирования.

3.7 **фактографическое индексирование:** Индексирование, предусматривающее отражение в поисковом образе документа конкретных сведений (сообщений), являющихся смыслом данного документа.

4 Общие положения

4.1 Процесс индексирования включает в себя следующие этапы, которые осуществляют в указанной ниже последовательности:

- анализ и определение содержания документа как объекта индексирования;
- выделение в документе тем, подлежащих отражению в виде отдельных тематических фрагментов;
- выбор понятий, характеризующих темы документа;
- выбор терминов индексирования для обозначения понятий;
- формирование поискового образа документа из терминов индексирования.

Перечисленные этапы могут быть объединены в составе технологических процедур при условии надлежащего выполнения каждого из этапов.

4.2 Поисковый образ документа (ПОД) формируют из выбранных терминов индексирования с помощью грамматических средств информационно-поискового языка (ИПЯ).

4.3 В процессе индексирования не рекомендуется описывать документ как физический объект (с точки зрения его формы, объема и пр.). Допускается отражать в ПОД подобную информацию, если она позволяет более точно установить соответствие документа информационной потребности пользователя системы.

5 Анализ документа

5.1 При анализе документа индексатору должна быть предоставлена возможность ознакомиться с документом в полном объеме. При невозможности исчерпывающего ознакомления с документом индексатор должен изучить имеющиеся текстовые части документа (основные источники индексирования):

- справочный аппарат документа — заглавие (наименование), аннотацию, реферат, содержание (оглавление), предисловие, заключение и др.;
- введение;
- заголовки частей и глав;
- первые абзацы глав и параграфов;
- иллюстрации, схемы, таблицы и подписи к ним;

- слова и группы слов, которые в тексте подчеркнуты или выделены полиграфическими средствами.

Индексирование только по заглавиям является неполноценным. При индексировании по рефератам и аннотациям следует следить за адекватностью передачи в них содержания документа.

5.2 При анализе нетекстовых (аудиовизуальных и других) документов, которые, помимо чтения, требуют просмотра, прослушивания, испытания объекта в действии и других подобных процедур, допускается индексирование их по имеющемуся текстовому компоненту (наименованию, краткому описанию и т.п.), но и в этом случае индексатору должна быть предоставлена возможность полного ознакомления с документом, если текстовой материал представляется недостаточным.

6 Выбор понятий, характеризующих содержание документа

6.1 Число характеристик и понятий, отраженных в ПОД, определяет его полноту и является важнейшим показателем качества индексирования.

6.1.1 В ПОД необходимо отразить все понятия всех тем, которые могут иметь ценность для пользователей системы.

В документе может быть выявлено более одной темы из сферы интересов пользователей. Эти темы должны рассматриваться раздельно.

6.1.2 Тематика, отражаемая при индексировании, не должна ограничиваться узкими рамками непосредственных интересов пользователей информационно-поисковой системы (ИПС). В ПОД также следует включать понятия, связанные с побочными аспектами документа (например, социальные и экономические аспекты научно-технических исследований).

6.1.3 При выборе понятий основным критерием является потенциальная ценность понятия для отражения темы документа или для его поиска. При этом необходимо ориентироваться на типичные запросы к ИПС:

- отбирать понятия, наиболее употребительные в коллективе пользователей ИПС;
- уточнять состав лексики и грамматические правила ИПЯ на основе обратной связи с пользователями.

Изменения, вносимые в ИПЯ, не должны нарушать общую структуру и логику, заложенные при его создании.

6.1.4 Число терминов индексирования, приписываемых одному документу, определяется количеством сведений, содержащихся в документе. Ограничение числа терминов должно быть основано на содержательном отборе наиболее важных понятий.

6.2 Полнота индексирования, принятая в каждой ИПС, определяется ее функциональным назначением. Объем документа сильно влияет на полноту индексирования. Необходимо учитывать указанные факторы и на их основе проводить экспертный отбор понятий из документа, не стремясь включить в ПОД все упомянутые в нем понятия.

6.3 Специфичность ПОД определяется тем, в какой мере понятия документа нашли точное отражение терминами индексирования, и также является одним из параметров качества индексирования. Представление понятия термином, имеющим более широкое значение, приводит к потере специфичности. Более широкие термины допускается использовать в случаях:

- если излишне специфичный термин непонятен пользователям, особенно когда соответствующее понятие применяется только в пограничных областях деятельности;
- если в документе понятие раскрыто недостаточно полно или является вспомогательным для изложения содержания документа.

6.4 Рекомендуется в каждой ИПС разрабатывать списки характеристик, которые признаются важными для отражения в ПОД. В зависимости от потребности конкретной ИПС этот список может быть как расширен, так и сокращен.

7 Выбор терминов индексирования

7.1 В процессе выбора терминов индексирования понятия, характеризующие содержание документа, представляют:

- предпочтительными лексическими единицами (дескрипторами или ключевыми словами), выбранными по правилам конкретного ИПЯ;

- терминами, отражающими новые понятия, проверив их точность по словарям, энциклопедиям, справочникам, классификационным таблицам, информационно-поисковым тезаурусам, терминологическим стандартам и другим источникам, признанным авторитетными и в данной области.

7.2 Выбор терминов индексирования осуществляют согласно *ГОСТ 7.25* или опубликованного информационно-поискового тезауруса, который используют при составлении запросов к ИПС.

При использовании тезауруса допускается сокращать число терминов, включаемых в ПОД, за счет исключения общих понятий, которые могут быть привлечены на этапе поиска документа или на этапе составления поискового предписания на основании ссылок в статьях тезауруса.

7.3 Понятия, не представленные в словаре индексирования, но необходимые для формирования ПОД, выражают одним из двух способов:

- новым специфическим термином, который включают в ПОД и в словарь;
- более общим термином, имеющимся в ИПЯ; при этом специфический термин направляют в службу ведения ИПЯ для включения в словарь.

Новые понятия представляют наиболее близкими из существующих в ИПЯ лексических единиц, а также оценивают полезность включения новых терминов в словарь с точки зрения эффективности поиска.

7.4 При индексировании свободными ключевыми словами, взятыми из текста документа, они должны быть приведены к канонической форме по *ГОСТ 7.25*. Длину словосочетаний рекомендуется ограничивать двумя-тремя словоформами.

Схема индексирования с использованием информационно-поискового тезауруса приведена в приложении А.

8 Формирование поискового образа документа

8.1 ПОД состоит из выбранных терминов индексирования, организованных с помощью грамматических средств ИПЯ данной ИПС.

8.2 В состав ПОД могут быть включены следующие категории данных, предусмотренные технологией индексирования конкретной ИПС:

- данные о степени нормализации терминов индексирования и применяемый для этого словарь;
- индивидуальные характеристики термина индексирования;
- связь терминов индексирования в синтаксических конструкциях ПОД.

Для включения в ПОД фактографических данных применяют грамматические категории, приведенные в разделе 9.

8.3 По степени нормализации различают два типа терминов координатного индексирования — дескрипторы и ключевые слова.

8.4 Термины индексирования должны быть представлены в ПОД в соответствии с орфографическими правилами используемого в системе естественного языка.

8.4.1 Дескрипторы допускается представлять условными кодами, которые указаны в используемом словаре индексирования. В этом случае ИПС должна обеспечивать автоматический поиск орфографических форм дескрипторов по их кодам.

8.4.2 Ключевые слова в многоязычных информационных системах с ПОД на основе различных национальных языков должны быть снабжены пометами о принадлежности к тому или иному естественному языку.

8.5 Индивидуальные характеристики терминов индексирования являются факультативными элементами ПОД, и их используют для уточнения содержания документа, организации процедур информационного поиска или дальнейшей аналитико-синтетической обработки документов в системе.

К индивидуальным характеристикам относят данные о семантической и морфологической категориях термина индексирования, его роли и информационном весе, способе получения и предполагаемом использовании.

8.5.1 Семантическая характеристика термина индексирования заключается в отнесении его к следующим лексикографическим категориям:

1) термин, выражающий научное или техническое понятие;
2) имя собственное, идентификатор;
3) наименование параметра;
4) значение параметра (выраженное текстом или именованной величиной);
5) числовое выражение;
6) обозначение единицы величины.

8.5.2 Морфологическая характеристика термина индексирования заключается в отнесении его к лексикографическим категориям:

1) производному слову;

2) сложному слову;

3) словосочетанию;

4) аббревиатуре;

5) фрагменту слова.

Морфологические характеристики используют в ПОД для реализации в ИПС смыслового анализа лексических единиц на основе их формальных признаков.

8.5.3 Роль термина индексирования указывают в ПОД для уточнения места соответствующего понятия в содержании документа. Для этого особыми указателями роли, принятыми в ИПС, отмечают термины индексирования, отражающие следующие аспекты документа:

а) объект исследования, описания;

б) характеристики, свойства, параметры объекта;

в) методы и средства исследования, технологическую оснастку,

г) составные части, узлы, детали объекта;

д) область применения объекта (отрасль хозяйства, техники, науки);

е) назначение объекта;

ж) цель исследования, разработки, описания;

з) результаты исследования, разработки.

8.5.4 Информационный вес термина индексирования отражает в ПОД важность данного понятия для данного документа. Число градаций информационного веса определяется потребностями конкретной ИПС. Следует различать понятия:

- выражающие главную тему документа;
- выражающие побочные темы документа;
- использованные в документе как вспомогательные для изложения его содержания.

Допускается использовать указатель отрицательного веса, которым помечают термины индексирования для указания на то, что данное понятие не рассматривается в документе.

8.5.5 Пометы, необходимые для указания на способ получения термина индексирования, используют для организации технологического процесса индексирования. Следует различать следующие пометы:

а) термин назначен по усмотрению индексатора, но отсутствует в документе;

б) термин введен в ПОД на основании связей, указанных в тезаурусе, но отсутствует в документе;

в) термин получен при автоматическом индексировании.

8.5.6 Пометы о предполагаемом использовании термина индексирования вводят в ПОД с целью выделить лексические единицы, подлежащие специальной обработке в процессах дальнейшей аналитико-синтетической переработки информации. Следует различать следующие пометы:

а) термин используется как предметная рубрика указателей;

б) при данном термине индексирования имеются фактографические данные, указанные в ПОД;

в) термин используется только как уточняющий определитель к другим терминам.

8.6 Термины индексирования в ПОД могут быть снабжены указателями связи, объединяющими их в синтаксические конструкции, которые отражают:

а) порядок следования и взаимное расположение терминов индексирования в документе;

б) смысловые связи понятий в документе;

в) парадигматические связи дескрипторов в тезаурусе.

Синтаксические конструкции рассматривают как цельные единицы ПОД наряду с терминами индексирования. Они могут быть объединены с другими синтаксическими конструкциями или с отдельными терминами индексирования в конструкции более высокого порядка.

Число уровней иерархии синтаксических конструкций определяется потребностями конкретных ИПС. Не следует применять конструкции четвертого и более высоких порядков.

Синтаксические конструкции могут быть охарактеризованы указателями веса, роли и предполагаемого использования аналогично индивидуальным терминам индексирования (см. 8.5.3, 8.5.4, 8.5.6).

8.7 Запись ПОД в памяти ИПС обусловлена принятым в ней способом кодирования с учетом требований настоящего раздела и *ГОСТ Р 7.0.52*.

9 Фактографическое индексирование документа

9.1 Фактографическое индексирование документа (ФИД) заключается в выявлении в документе и включении в ПОД данных, выражающих конкретные сведения (сообщения), имеющиеся в документе.

На основании результатов ФИД в фактографических ИПС формируются массивы сведений, в которых единицей информации является фактографическая запись.

9.2 ФИД предполагает формальное различение в ПОД двух категорий терминов индексирования, выражающих:

а) темы или объекты сообщения;

б) приписанные этим объектам свойства, являющиеся смыслом сообщения.

Соответствующие термины индексирования должны быть связаны друг с другом в синтаксическую конструкцию, объединяющую наименование объекта, его характеристики, их значения, единицы величины и отражающую смысловые связи понятий в документе.

Дополнительно такая синтаксическая конструкция может быть охарактеризована:

- показателем модальности;
- условием истинности.

9.3 Показатель модальности фактографического сообщения определяет различие между сообщениями следующих типов:

- наблюдаемый факт;
- допустимое значение;
- требование стандарта;
- плановый показатель;
- запрет,
- рекомендация;
- предположение;
- условие.

Если в информационной системе не используют показатели модальности, то все фактографические сообщения рассматривают как принадлежащие одной модальности, которая должна быть указана в эксплуатационной документации системы.

9.4 Условием истинности фактографического сообщения является другое фактографическое сообщение, связанное с первым в синтаксическую конструкцию вышестоящего уровня.

Пример

X = вес продукта (X — характеристика объекта);
Z = 150 г (Z — значение характеристики);
Y = влажность не более 45 % (Y — условие истинности).

Фактографическое сообщение, являющееся условием истинности, должно иметь показатель модальности условия «если», например:

(вес продукта = 150 г) (если влажность не более 45 %).

9.5 Термины индексирования, выражающие тему (объект) сообщения, относятся к категории 1 или 2, указанной в 8.5.1. При использовании категории 1 термину индексирования может быть дополнительно приписан показатель единичности или общности объекта (квантор).

Квантор общности используют в сообщениях, где выражено утверждение обо всех объектах, попадающих в объем соответствующего понятия.

Квантор единичности используют в сообщениях, где выражена информация о том объекте, входящем в состав данного понятия, который рассматривается в данном документе.

9.6 Термины индексирования, выражающие свойства объектов, которые составляют смысл сообщения, могут быть выражены лексическими единицами категорий 1, 2, 3 (см. 8.5.1) или параметрической конструкцией (см. 8.6).

9.7 Параметрическая конструкция должна состоять из двух формально выраженных частей: наименования параметра и перечня значений параметра (см. 9.8), которые объединены в одну синтаксическую конструкцию.

9.8 Перечень значений в параметрической конструкции должен включать в себя набор значений параметров и указание об альтернативности или одновременности (симультанности) значений.

Набор значений задают перечислением или указанием двух предельных значений, между которыми располагаются значения, принимаемые параметром (интервалом значений). При задании интервала значений формально указывают, которое из значений является начальным и конечным для интервала

значений, а также, входят ли граничные значения в указанный интервал. Одно из граничных значений интервала может отсутствовать, если значение параметра ограничено только с одной стороны.

Указание об одновременности используют, когда у одного объекта сообщения наблюдаются все заданные значения параметра. Указание об альтернативности используют, когда параметры одного объекта сообщения должны быть выбраны из числа заданных.

9.9 Значения параметра могут быть представлены синтаксической конструкцией из двух терминов индексирования — числового выражения и наименования единицы величины — при необходимости производить операции расчета или численного сравнения.

9.10 Числовые значения и их размерности в ПОД должны соответствовать требованиям к представлению числовых данных.

10 Автоматизированное индексирование

10.1 Цель автоматизации индексирования — это минимизация материальных и человеческих ресурсов, затрачиваемых на процедуру индексирования, а также достижение стабильности и единообразия ее результатов.

10.2 Автоматизированное индексирование (АИ) осуществляют:

а) по тексту первичного документа;

б) по заглавию и аннотации или реферату документа.

АИ по тексту первичного документа должно включать в себя процедуру сжатия ПОД.

10.3 С использованием компьютерной техники осуществляют следующие содержательные этапы АИ:

1) выявление информативных частей документа;

2) идентификацию слов текста и приведение их к нормализованному виду (морфологический анализ и синтез);

3) формирование списка ключевых слов исходного текста;

4) подбор дескрипторов по тезаурусу;

5) формирование ПОД.

10.4 Выявление информативных частей документа

Технология АИ должна предусматривать идентификацию и предоставление индексатору или программе индексирования наиболее информативных фрагментов документа из списка, указанных в 5.1. Могут быть предусмотрены алгоритмы выявления информативных фрагментов по другим формальным критериям, а также по решению специалиста-индексатора.

10.5 Идентификация слов текста

10.5.1 Процесс идентификации слов текста должен включать в себя отождествление словоформ одного слова и определение ключевых слов текста. При этом может быть необходимо использование интеллектуальных процедур для решения таких задач, как выявление и обработка синтаксических конструкций, выявление и разрешение омонимии.

10.5.2 Для идентификации слов текста используют машинные словари (словари основ, парадигм, словосочетаний и т.д.). Словари должны быть представлены в базе данных системы и обеспечены средствами визуализации и ведения.

10.6 Формирование списка ключевых слов текста

10.6.1 В процессе формирования списка ключевых слов текста проводят синтаксический анализ текста с учетом правил сочетаемости грамматических категорий данного естественного языка.

10.6.2 Синтаксический анализ текста решает задачи:

а) разделение текста на фрагменты по заданным критериям;

б) установление синтаксических зависимостей между словоформами текста;

в) отождествление словосочетаний;

г) нормализация выявленных ключевых слов.

10.7 Автоматическое формирование ПОД

10.7.1 В процедуре АИ допускается формирование ПОД из свободных ключевых слов или дескрипторов информационно-поискового тезауруса, используемого в данной области.

10.7.2 При АИ дескрипторами информационно-поискового тезауруса на этапе формирования ПОД происходит замена ключевых слов на дескрипторы, указанные в тезаурусе.

10.7.3 При формировании ПОД из дескрипторов рекомендуется обогащать ПОД за счет пополнения вышестоящими терминами информационно-поискового тезауруса.

10.7.4 Процедура АИ должна предусматривать включение в ПОД типовых грамматических средств (см. раздел 8).

10.7.5 К системам АИ предъявляются следующие требования:

- модульность построения (внутренняя организация лингвистического и программного обеспечений системы, при которой процедуры решения отдельных задач АИ реализуются с помощью самостоятельных блоков или модулей);
- ориентация на типовые программные и технические средства;
- соответствие действующим нормативно-методическим документам по координатному индексированию.

Приложение А
(справочное)

Схема индексирования по информационно-поисковому тезаурусу

А.1 Изучить документ и составить перечень существенных для его содержания понятий с учетом специфики ИПС.

А.2 Рассмотреть первое понятие.

А.3 Найти в тезаурусе лексическую единицу, отражающую данное понятие. Если таковой нет, то перейти к А.11.

А.4 Если найденная лексическая единица — аскриптор, то заменить ее указанным в ссылке дескриптором (или комбинацией дескрипторов).

А.5 Рассмотреть ссылки, указанные в тезаурусе для данного дескриптора (дескрипторов).

А.6 Проверить, не являются ли указанные в ссылках дескрипторы более специфичными для выражения данного понятия. Если да, то перейти к А.10.

А.7 Записать найденные лексические единицы в поисковый образ, снабдив их необходимыми грамматическими показателями по правилам данного ИПЯ.

А.8 Проверить, имеются ли еще не отраженные в поисковом образе понятия из документа, и рассмотреть следующее понятие. Перейти к А.3.

А.9 Если список понятий документа исчерпан, то окончить работу.

А.10 Заменить исходный дескриптор более специфичными согласно указанию ссылки в тезаурусе. Перейти к А.7.

А.11 Найти в тезаурусе дескрипторы, совместное включение которых в поисковый образ отражает данное понятие. Если таковых нет, то перейти к А.12, если есть — к А.5.

А.12 Установить термин, выражающий понятие и удовлетворяющий требованиям к дескрипторам по *ГОСТ 7.25*.

А.13 Направить найденный термин в службу ведения ИПЯ в качестве кандидата на включение в тезаурус. Перейти к выполнению А.7.

УДК 389.6:002:006.354 ОКС 01.140.40

Ключевые слова: индексирование, координатное индексирование, информационно-поисковый тезаурус, лексическая единица, информационно-поисковый язык, ключевое слово, дескриптор

Редактор *Н.В. Таланова*
Технический редактор *И.Е. Черепкова*
Корректор *С.В. Смирнова*
Компьютерная верстка *А.А. Ворониной*

Сдано в набор 28.12.2018. Подписано в печать 16.01.2019. Формат 60×84 $^{1}/_{8}$. Гарнитура Ариал.
Усл. печ. л. 1,86. Уч.-изд. л. 1,15.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

ИД «Юриспруденция», 115419, Москва, ул. Орджоникидзе, 11.
www.jurisizdat.ru y-book@mail.ru

Создано в единичном исполнении ФГУП «СТАНДАРТИНФОРМ» для комплектования Федерального информационного фонда стандартов 117418 Москва, Нахимовский пр-т, д. 31, к. 2.
www.gostinfo.ru gost.e-ecolog.ru

gost.e-ecolog.ru