ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р ИСО 23185— 2013

ОЦЕНКА И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТЕРМИНОЛОГИЧЕСКИХ РЕСУРСОВ

Общие понятия, принципы и требования

ISO 23185:2009
Assessment and benchmarking of terminological resources — General concepts, principles and requirements
(IDT)

Издание официальное

Москва
Стандартинформ
2015

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН ЗАО «Проспект» на основе собственного аутентичного перевода на русский язык международного стандарта, указанного в пункте 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 55 «Терминология, элементы данных и документация в бизнес-процессах и электронной торговле»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 8 ноября 2013 г. № 1390-ст

4 Настоящий стандарт идентичен международному стандарту ИСО 23185:2009 «Оценка и сравнительный анализ терминологических ресурсов. Общие понятия, принципы и требования» (ISO 23185:2009 «Assessment and benchmarking of terminological resources — General concepts, principles and requirements»).

При применении настоящего стандарта рекомендуется использовать вместо ссылочных международных стандартов соответствующие им национальные стандарты Российской Федерации, сведения о которых приведены в дополнительном приложении ДА

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, уведомления 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, 2015

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 Использование терминологических ресурсов . . . 2
4 Терминологические ресурсы . . . 2
4.1 Общие положения . . . 2
4.2 Оценочная модель для терминологических ресурсов . . . 3
4.3 Общие характеристики полезности терминологических ресурсов . . . 4
5 Руководящие принципы оценки и сравнительного анализа . . . 10
5.1 Общие положения . . . 10
5.2 Основные этапы работы . . . 10
5.3 Тестовая оценка . . . 10
Приложение А (обязательное) Характеристики полезности терминологических ресурсов и основные правила их оценки . . . 11
Приложение В (справочное) Типовое использование характеристик полезности как эталонов . . . 15
Приложение ДА (справочное) Сведения о соответствии ссылочных международных стандартов национальным стандартам Российской Федерации . . . 18
Библиография . . . 19

Введение

Под влиянием научно-технического прогресса мировое сообщество все быстрее движется в своем развитии по пути формирования многоязычного информационного общества знаний, для которого характерно все более глубокое проникновение информационно-коммуникационных технологий во все сферы жизни. Для поддержания высоких темпов развития индустрии знаний и информационных ресурсов крайне важно иметь надежные языковые ресурсы (такие как собрания текстов и речей, терминология, словари прикладной лингвистики и др.). Таким образом, терминологическая информация становится ключевым элементом любых регуляционных действий, которые, например, хорошо заметны в области технической стандартизации, управлении качеством и охране прав интеллектуальной собственности.

В постоянно расширяющейся семантической сети специализированные и неспециализированные браузеры или сетевые службы обеспечивают поиск сетевых баз данных и порталов со структурированным содержимым (т.е. совокупностью информационных элементов на уровне лексической семантики). Пользователь в ответ на свой запрос хочет иметь максимально релевантные и надежные сведения, в которых не потеряна важная для него информация. Поэтому для удовлетворения реальных пользовательских потребностей результаты автоматического или полуавтоматического поиска все больше нуждаются в соответствующей компоновке и сжатии с применением методов семантического анализа.

Если в результате поиска по массивам структурированного контента отыскивается информация, кажущаяся релевантной, то далее должны выполняться систематические процедуры синтаксической и семантической фильтрации, отбора и оценивания. На каком-то этапе этих процедур браузеры или сетевые службы должны осуществлять четкое выделение из всего массива информации релевантных и достоверных терминологических данных, которые собираются в целях установления соответствующих приоритетов и оптимизации.

Терминологические данные могут служить множеству разных целей, из которых наиважнейшими являются:

- представление знаний (формулирование концептуальных понятий);
- упорядочение системы знаний (классификация понятий);
- обеспечение доступа к другим типам структурированного либо неструктурированного контента и
- предоставление средств или элементов коммуникации и передачи знаний.

Применительно к таким сферам, как перевод, локализация и управление контентом, оказываются необходимыми также систематические методы автоматической или полуавтоматической оценки и сравнительного анализа ресурсов или собранных терминологических данных.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ОЦЕНКА И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТЕРМИНОЛОГИЧЕСКИХ РЕСУРСОВ

Общие понятия, принципы и требования

Assessment and benchmarking of terminological resources.
General concepts, principles and requirements

Дата введения — 2015—01—01

1 Область применения

В настоящем стандарте описываются основополагающие принципы эффективного использования терминологических данных; излагаются общие принципы построения модели, применимой к самым разным терминологическим ресурсам, дается определение атрибутов применимости, заложенных в структуру модели, и формулируются руководящие принципы глобальной оценки терминологических ресурсов с учетом специфики пользовательских задач.

2 Термины и определения

В настоящем стандарте используются следующие термины с соответствующими определениями.

2.1 **оценка** (assessment): <*в терминологии*> Контрольный процесс, призванный продемонстрировать соответствие некоторого терминологического ресурса (2.8) конкретным требованиям.

2.2 **эталон** (для сравнения) (benchmark): <*в терминологии*> Значение характеристики полезности (2.11), используемое в качестве отправного пункта или начала отсчета, относительно которого может измеряться полезность (2.10) конкретного терминологического ресурса (2.8).

2.3 **сравнение с эталоном** (benchmarking): <*в терминологии*> Использование эталонов (2.2) применительно к терминологическим ресурсам (2.8).

2.4 **сущность** (entity): Любой конкретный или абстрактный предмет, реально существующий или который мог бы существовать вместе с его возможными связями.

Пример — Индивидуум, объект, событие, идея, процесс и т. п.

[ИСО/МЭК 2382-17:1999, 17.02.05]

2.5 **оценочная модель** (model for assessment): <*в терминологии*> Модель, идентифицирующая характеристики полезности (2.11) терминологических ресурсов (2.8) и их взаимосвязи.

2.6 **специальный язык** (special language): Язык, используемый в рамках предметной области и характеризуемый наличием специфических лингвистических выразительных средств.

[ИСО 1087-1:2000, определение 3.1.3]

2.7 **терминологические данные** (terminological data): Данные, относящиеся к понятиям или их обозначениям.

[ИСО 1087-1:2000, определение 3.8.1]

2.8 **терминологический (информационный) ресурс** (terminological resource; terminological data resource): Сущность (2.4), образованная совокупностью терминологических данных (2.7) с характеристиками полезности (2.11), которые созданы в результате группировки / структурирования данных или их интегрирования в приложение.

Издание официальное

Примечание 1 — Терминологический ресурс обычно содержит терминологические данные (2.7), которые структурированы (как, например, терминологическая база данных), размечены с помощью языка разметки (например, представлены в виде информационного XML-файла) или ассоциируются с некоторым структурированным набором терминов (например, со словарем). Терминологический ресурс может содержать простые тексты (например, тексты разных стилей одного языка), из которых могут быть извлечены полезные терминологические данные с помощью какой-либо современной технологии обработки данных.

Примечание 2 — Обычно терминологические ресурсы, создаваемые как продукт профессиональной терминологической деятельности, бывают хорошо организованы и структурированы и содержат высококачественные терминологические данные (2.7), благодаря которым эти ресурсы обладают достаточно высокими характеристиками полезности (2.11). В отличие от этого, терминологические ресурсы, создаваемые в рамках других процессов и действий, часто характеризуются непредсказуемыми показателями качества и полезности.

2.9 **терминология** (terminology): Система символических обозначений, применимая к понятиям, которые относятся к одному специальному языку (2.6).

Примечание — Определение, данное в ИСО 1087-1, выглядит несколько иначе: «система символических обозначений, принадлежащих одному специализированному языку» (ИСО 1087-1:2000, определение 3.5.1).

2.10 **полезность** (usability): Степень пригодности некоторой сущности для эффективного и продуктивного достижения определенных целей приемлемым способом.

2.11 **характеристика полезности** (usability attribute): <*в терминологии*> Атрибут терминологических данных (2.7), по которому определяется полезность (2.10).

Примечание 1 — Характеристики полезности могут быть измеримыми или неизмеримыми. Измеримые характеристики полезности могут оцениваться количественно, тогда как неизмеримые подлежат качественному описанию.

Примечание 2 — Характеристика полезности описывается в соответствии с заданными требованиями или измеряется с использованием заданного метода.

3 Использование терминологических ресурсов

Как правило, терминологические ресурсы требуются для того, чтобы:

- использовать контент базы знаний, относящийся к понятиям или их представлениям, в специальном языке (например, путем поиска терминов в словаре, обращения с запросом к базе данных, информационного поиска на терминологическом портале);
- управлять информацией, относящейся к конкретной предметной области (например, к системам электронного бизнеса, электронного правительства и дистанционного / онлайн обучения);
- организовывать эффективный обмен терминологическими данными (например, между различными сферами деятельности, информационными системами и терминологическими ассоциациями; импортировать терминологические данные в системы переводческой памяти (translation memory); экспортировать терминологические данные для составления специализированных словарей);
- облегчать терминологическую работу (например, для специалистов по терминологии);
- объединять воедино рассредоточенные и дублируемые терминологические ресурсы для обеспечения эффективной работы информационных служб (например, оперативно доступных служб, которые обеспечивают выдачу специализированной информации, ввод и управление данными кооперативных вычислений и реализуют технологии семантической web-паутины).

4 Терминологические ресурсы

4.1 Общие положения

Понятие «терминологические ресурсы» должно рассматриваться во всем его многообразии:

a) Терминологическая информация представляется, записывается или сохраняется на информационных носителях как результат процессов подготовки, регистрации и обработки данных.

b) Для использования терминологических данных требуется выполнение таких операций, как:

1) обработка с применением самых разных технологий с целью предоставления пользователям необходимых услуг;
2) запрос соответствующего логического контента, эпистемологических форм (относящихся к теории познания) и лингвистических выражений области знаний, представляемой терминологическими данными;

3) передача или преобразование специальных знаний для интегрирования их в пользовательский языковой контекст, например в процессе перевода или локализации; однако проблема определения соответствия или несоответствия терминологических данных целевому языковому контексту выходит за рамки настоящего стандарта.

c) С точки зрения пользователя, терминологические данные служат двум основным целям:

1) они удовлетворяют потребности пользователя в получении нужного ему контента терминологического характера и

2) предназначаются для обеспечения возможности выполнения желаемых операций доступа к данным или их обработки.

Отсюда следует, что при анализе терминологических ресурсов речь идет, как правило, о применении к ним тех или иных информационных технологий, а это значит, что для превращения терминологических данных в используемый терминологический ресурс обычно требуются определенные средства их технологической обработки.

Терминологический ресурс должен рассматриваться как динамическая интегрированная система терминологических данных. Как только представлены терминологические данные, эта система начинает существовать, развиваться и выполнять свои функции. Эффективность использования (полезность) этой системы отражается ее характеристиками полезности. Эти характеристики четко определяются в настоящем стандарте.

Терминологические ресурсы могут подвергаться систематической оценке. При наличии соответствующих инструментальных средств такая оценка может выполняться автоматически. Терминологические ресурсы подлежат оценке на основе анализа механизмов их создания, возможностей их использования и способов управления ими.

Пример 1 — Словарь на цифровом носителе, содержащий тысячу терминологических статей, обычно снабжается руководством пользователя, в котором объясняется, как осуществляется доступ к носителю, какое оборудование для этого требуется и какие инструментальные средства нужны для считывания данных. Пользователи в этом случае могут, следуя указаниям руководства, просматривать данные и определять, пригодны ли они для их конкретных целей. Такой терминологический ресурс состоит из ревльной информации и ее носителя, которые доопределяют характеристики полезности, приведенные в руководстве пользователя.

Пример 2 — Словарь терминов по математике, физике и химии тоже является терминологическим ресурсом. Теоретически для пользователя, которому нужен доступ только к химической терминологии, этот словарь будет мало отличаться от словаря, содержащего лишь терминологию по химии, то есть от словаря без математических или физических терминов. Однако совершенно иначе будет оцениваться полезность такого словаря пользователями, которые обращаются к нему за математической или физической терминологией.

В двух приведенных выше примерах демонстрируется методика оценки терминологических данных, при которой центром внимания являются:

- способ группирования терминологических данных с учетом их ассоциативных связей как терминологического ресурса;
- оценка характеристик полезности терминологического ресурса на основе всестороннего анализа следующих элементов: способа записи данных, способа хранения данных, формата данных, структуры данных, обозначенной надлежащим образом предметной области, практических потребностей пользователя и т. п.

4.2 Оценочная модель для терминологических ресурсов

Модель для оценки терминологических ресурсов должна включать в себя четыре группы характеристик полезности, относящихся к следующим компонентам:

- терминологические данные;
- схема управления данными;
- вывод данных;
- ввод данных.

Не обязательно, чтобы каждая группа характеристик полезности соответствовала каждому терминологическому ресурсу.

Пример 1 — Если печатный экземпляр специализированного словаря считается терминологическим ресурсом, то он не должен снабжаться характеристиками полезности ввода данных и управления данными; его характеристики полезности должны включать только оценки, касающиеся выходной информации (печатных страниц), а именно — присутствующих на этих страницах терминологических данных и указателей.

Опять же не обязательно, чтобы каждая группа характеристик полезности, относящихся к способу управления данными, выводу и вводу данных, была присуща каждому терминологическому ресурсу.

Пример 2 — Если речь идет о базе данных, то отображение ее индекса в пользовательских интерфейсах не обязательно. Пользователь обычно не считает такой отображаемый индекс атрибутом, полезным для практического использования.

Приведенный ниже рисунок 1 иллюстрирует общую структуру оценочной модели.



Рисунок 1 — Четыре группы характеристик полезности

Хотя в последующих разделах приводится индивидуальное описание каждого атрибута, необходимо иметь в виду, что характеристики полезности могут быть взаимосвязаны и могут зависеть друг от друга.

4.3 Общие характеристики полезности терминологических ресурсов

4.3.1 Характеристики полезности, относящиеся к терминологическим данным

Терминологический ресурс должен содержать определенное число атрибутов, которое достаточно для представления сложной структуры терминологических данных и в то же время соответствует конкретным целевым требованиям; такими атрибутами должны быть:

- спецификация структуры данных;
- охват категорий данных;
- охват предметных областей;
- охват языков;
- признак соответствия правилам смысловой связности;
- признак использования внешних контрольных данных;
- ссылки на используемые авторитетные источники;
- указание обладателя прав интеллектуальной собственности;
- симметрия представления коллекции терминологических данных;
- объем терминологических данных.

4.3.1.1 Спецификация структуры данных

Спецификация структуры данных касается модели данных терминологического ресурса. Понимание модели данных позволяет пользователям эффективно применять надлежащие технологии извлечения терминологических данных из терминологического ресурса, производить настройку служб, использующих терминологический ресурс, и т. п. Отсюда следует, что спецификация структуры данных должна быть предельно понятна для пользователей, и это является правилом для данной характеристики полезности.

4.3.1.2 Охват категорий данных

Этот атрибут указывает множество категорий, используемых в модели терминологического ресурса. Состав категорий данных изменяется в зависимости от предметных областей или сфер применения.

Пример — ИСО 12616 устанавливает категории данных, необходимые для терминографии, ориентированной на выполнение переводов. Это определяет метрику, которая позволяет выяснить, подходит ли данный терминологический ресурс для переводческой терминографии.

ИСО 12620 определяет возможные категории для регистрации терминологических данных.

Атрибут «охват категорий данных» измерим, если используются категории, определенные ИСО 12620, и если модель данных строится в соответствии с требованиями ИСО 16642. Если же модель данных и категории данных терминологического ресурса не соответствуют ИСО 16642 и ИСО 12620, то для оценки терминологического ресурса необходим механизм преобразования.

4.3.1.3 Охват предметных областей

Атрибут «охват предметных областей» относится к группе предметных областей, знания которых представляются терминологическими данными терминологического ресурса.

Терминология всегда связывается с конкретным языком в определенной области знаний. Эта область знаний должна представляться одним или несколькими указателями предметных областей, относящихся:

- к определенной научной дисциплине в установленной схеме классификации либо в контролируемом языке, или
- к конкретной сфере применения терминологии либо к конкретному сообществу специалистов предметной области.

Если такого разделения нет, то должны явным образом указываться конкретные предметные области. Указатель предметной области, который либо представлен на контролируемом языке, либо является обычно используемым явным указателем, может относиться к совокупности терминологических данных в целом, к каждой статье в отдельности или к тому и другому одновременно.

Охват предметных областей измерим, когда контролируемые языки или явные указатели предметной области, используемые для ее обозначения, сопоставимы. Впоследствии пользователи могут сравнить предметные области, фактически покрываемые терминологическим ресурсом, со своими ожиданиями.

4.3.1.4 Охват языков

Полнота представления языков определяется идентификаторами языков, то есть набором названий или символов, обозначающих естественные языки либо их варианты. Этот атрибут показывает один или несколько языков, используемых для записи терминологических данных, и указывает, сколько языков охватывает информация терминологического ресурса.

Указатель языка терминологического ресурса должен анализироваться в следующих аспектах:

- языковые разделы;
- язык описания метаданных;
- география применения;
- системы записи и системы кодирования символов.

Стандартами серии ИСО 639 устанавливаются коды для представления названий языков. В ИСО 15924 предлагается стандартизированный код для представления гарнитур шрифтов. Указатели языков могут детализироваться путем добавления кодов, идентифицирующих географические особенности применения языка в соответствии с требованиями серии ИСО 3166.

Если указатель языка, присутствующий в терминологическом ресурсе, соответствует вышеуказанным стандартам, то характеристика полезности становится измеримой, то есть она может быть сопоставлена с атрибутами практической полезности без дополнительного преобразования.

4.3.1.5 Соответствие правилам когерентности

Соответствие правилам когерентности проверяется путем соответствующего анализа правил, установленных применительно к записываемым терминологическим данным. Эти правила обычно контролируют взаимозависимость между ними, то есть

- *перекрестные ссылки между терминами и лексическими единицами, в которых они определены:*
  - ссылочные связи могут проверяться на достоверность, отслеживаться и подвергаться контролю вручную либо автоматически;
  - независимо от того, правильные ли перекрестные ссылки или ошибочные, они могут быть оценены количественно;
- *использование унифицированной орфографии, унифицированной фразеологии и т. п. для представления познавательной информации:*
  - эта характеристика неизмерима;
- *организация расположения данных в соответствии с определенными принципами:*

- некоторые компьютерные программы способны формировать системы понятий на основе анализа существующих данных; такие программы могут помогать пользователям определить, содержит ли конкретный терминологический ресурс систематизированные терминологические данные; из этого следует, что этот атрибут вычисляется посредством качественной оценки, а не путем простого измерения.

Независимо от того, могут или не могут все вышеуказанные правила применяться в явной форме, терминологический ресурс должен быть оценен путем анализа всех рассмотренных выше аспектов.

4.3.1.6 Использование контролируемых внешних данных

Под использованием контролируемых внешних данных понимается обращение к дополнительной информации[1)], которая образует внешние ресурсы, такие как стандарты с независимым сопровождением, тезаурусы, онтологии и библиографии. Данный атрибут позволяет оценивать качество данных: чем качественнее приведенные ссылки, тем выше будет ожидаемое качество данных. Если в работе используется компьютер, то ссылки на внешние ресурсы обрабатываются автоматически и качество ресурсов может оцениваться количественно.

4.3.1.7 Использование авторитетных источников

Под использованием авторитетных источников понимается обращение к заслуживающим доверия документам или другим источникам, из которых берутся терминологические данные. Чем выше авторитет используемого источника, тем выше будет и качество получаемых данных. Критерии «авторитетности» источников носят качественный характер.

4.3.1.8 Указание обладателя прав интеллектуальной собственности

Под указанием обладателя прав интеллектуальной собственности понимается указание владельца конкретных данных в лице соответствующих индивидов, организаций, учреждений или институтов. Идентификатор правообладателя часто присоединяется к названию источника информации. Косвенно такой идентификатор предупреждает пользователей о том, что владелец прав установил определенные ограничения, которые подлежат обязательному выполнению при первичном или повторном использовании терминологических данных ресурса. Эта характеристика полезности неизмерима.

При более строгом подходе к охране авторских прав с применением информационных технологий подобные идентификаторы могут распространяться как на весь терминологический ресурс, так и на каждую его статью или даже на отдельные элементы информации. Если права интеллектуальной собственности (или разрешение на использование ресурса) описаны подробно и в контролируемой форме, то статус авторского права на терминологический ресурс может представляться некоторым качественным статистическим показателем, а это делает данную характеристику полезности измеримой.

4.3.1.9 Симметрия представления совокупности терминологических данных

Под симметрией представления совокупности терминологических данных понимается требование полного представления информации на разных языках. Эта характеристика полезности может оцениваться значениями «полная», «неполная» или числом полных информационных блоков и т. п.

Пример — Крайне важно, чтобы определение на каждом из охватываемых языков в двуязычном или многоязычном терминологическом ресурсе было полным и одинаково доступным для использования. Если в двуязычном терминологическом ресурсе полная информация представлена и доступна только применительно к одному языку, то такой ресурс не должен оцениваться как двуязычный.

4.3.1.10 Объем коллекции терминологических данных

Объем коллекции терминологических данных характеризует полный объем доступной информации терминологического ресурса. Этот объем может измеряться числом статей или количеством байтов, мегабайтов и др. Объем отдельного терминологического ресурса может быть разным в оценках, которые предназначены для разных целей, предметных областей или сфер применения. Эта характеристика служит эталоном для оценки степени полноты различных терминологических ресурсов.

Пример — Такими атрибутами могут быть число символов (по возможности — для каждого языка), количество определений и др.

4.3.2 Характеристики полезности, связанные с механизмом управления данными

Характеристики полезности, связанные с управлением данными, касаются методов и механизмов, которые должны использоваться в терминологическом ресурсе для следующих целей:

[1)] В ИСО 16642:2003, пункт 3.1, дополнительная информация определяется как «информация, дополняющая сведения, приведенные в терминологических статьях, и являющаяся общей для всей коллекции терминологических данных».

- проверки достоверности (валидации) данных:
  - полноты,
  - лингвистической корректности,
  - соответствия условиям представления;
- проверки на избыточность или дублирование;
- регулярному сопровождению данных;
- регулярному сопровождению метаданных (например, реестра метаданных в соответствии с требованиями ИСО 11179).

Многие из характеристик полезности, описанные в данном подразделе, связаны с атрибутами ввода или вывода данных.

4.3.2.1 Проверка достоверности данных

Проверка достоверности — это проверка системной корректности терминологических данных. Если пользователь ознакомлен с правилами контроля достоверности представляемых данных, это помогает ему определять их текущее состояние и даже делать это в автоматическом режиме.

Вследствие сложности структуры терминологических данных, вышеуказанная характеристика полезности должна анализироваться в разных аспектах: на полноту, лингвистическую корректность и на соответствие требованиям к представлению данных.

4.3.2.1.1 Полнота

Полнота означает уверенность в том, что записанные терминологические данные состоят из элементов, которые отвечают требованиям соответствующих категорий, определенных в модели данных. Этот атрибут поддается измерению, особенно в компьютеризованной рабочей среде.

Пример — В подразделе 6.1 ИСО 10241:1992 констатируется, что «для целей стандартизации терминологическая статья должна содержать хотя бы следующие элементы: a) номер статьи; b) предпочтительный термин, представляющий понятие; c) определение понятия...». Если терминологическая статья не отвечает этому минимальному требованию, то она не должна квалифицироваться как статья стандартизованной терминологии ИСО. В этом случае может использоваться некоторый статистический показатель, указывающий степень соответствия ИСО 10241:1992, подраздел 6.1.

4.3.2.1.2 Лингвистическая корректность

Этот атрибут относится к проверке лингвистической правильности терминологических данных. Для оценки лингвистической корректности необходимо проверить выполнение основных лингвистических соглашений и зависящих от предметной области ограничений, применимых к терминологическим данным. Эта характеристика измерима — особенно в компьютеризованной рабочей среде.

Пример — Для многих естественных языков существуют орфографические корректоры. Использование таких корректоров позволяет пользователю оценивать лингвистическую корректность с помощью статистики ошибок, выраженной в абсолютном их количестве или в процентах.

4.3.2.1.3 Соответствие условиям представления

Данный атрибут касается стандартизованного представления контента терминологического ресурса. Он особенно просто измеряется в компьютеризованной рабочей среде.

Пример 1 — Коды для представления названий языков, географических названий и др. должны быть стандартными, такими, как они определены в ИСО 639, ИСО 3166 и ИСО 15924. Соответствие настоящему стандарту может оцениваться с помощью статистических показателей.

Пример 2 — Представление терминологических данных в напечатанных стандартах ИСО по терминологии соответствует спецификациям ИСО 10241.

4.3.2.2 Проверка на избыточность

Под проверкой на избыточность понимается выявление случаев повторного появления одной и той же содержательной информации в терминологическом ресурсе. Установить все случаи дублирования информации в терминологии непросто, поскольку терминологические данные всегда привязаны к конкретным понятиям определенной предметной области.

Вопрос избыточности рассматривается в спецификациях ИСО 860, где на определение избыточности накладываются различные ограничения в соответствии со схемами классификации предметных областей. Важным фактором оценки избыточности может быть также структура механизма управления терминологическим ресурсом.

Обычно пользователи не хотят иметь дела с дублированным контентом, так как это замедляет работу с терминологическим ресурсом. Однако в определенных ситуациях значимой может оказаться характеристика полезности «Контроль за уровнем дублирования понятий».

Этот контроль касается ненужных повторных вхождений одних и тех же данных, привязанных к одному и тому же понятию в одном и том же терминологическом ресурсе.

Пример — В древовидной структуре системы предметных областей идентичные понятия не могут существовать в рамках одной предметной области, расположенной в конце ветви «дерева».

На практике критерии оценки дублирования понятий основываются на рассмотрении всего контента и структуры терминологического ресурса.

Применительно к терминологическому ресурсу это правило оценки применяется для обеспечения экономного описания атрибутов, то есть для использования наименьшего количества данных, что наилучшим образом соответствует практическим потребностям.

4.3.2.3 Регулярное сопровождение данных

Регулярное сопровождение данных заключается в обновлении терминологической информации терминологического ресурса в соответствии с последними разработками в определенной области знаний. Этот атрибут может измеряться посредством обработки доступных в терминологическом ресурсе сведений о дате. Однако для оценки значения этой характеристики часто требуется оценка познавательного контента: например, в тех случаях, когда вновь созданная запись содержит устаревшие сведения.

4.3.2.4 Регулярное сопровождение метаданных

Регулярное сопровождение метаданных — это поддержание соответствия информационной структуры и спецификации категорий данных реестрам метаданных в целях упрощения информационного обмена между терминологическими ресурсами и обеспечения более удобного повторного использования данных. Ключевым аспектом оценки этой характеристики является наличие поддержки удобного обмена данными и возможности их многократного использования. (См. также п. В.2.3 в приложении В).

4.3.3 Характеристики полезности, относящиеся к выводу данных

Характеристики полезности, относящиеся к выводу данных, касаются доступности терминологических данных. Они характеризуют непосредственные информационные нужды пользователей в следующих аспектах:

- читабельность выходных данных;
- доступ к данным;
- пользовательский отбор данных.

4.3.3.1 Читабельность выходных данных

Читабельность выходных данных определяется форматом, в котором терминологическая информация предоставляется пользователям.

Пример — В качестве примера можно назвать конфигурации диалоговых окон на экране компьютерного дисплея (пользовательские интерфейсы), схему расположения статей в словаре и т. п.

В ИСО/МЭК 9126 (все части) и ИСО 9241-11 определяются атрибуты технической полезности компьютерных интерфейсов; ИСО 10241 обусловливает спецификации шаблонов для разработки терминологических стандартов ИСО.

Читабельность выходных данных — неизмеримый атрибут. Наиболее важными факторами для описания этого атрибута являются:

- пользовательские предпочтения;
- выделение эффективной категории данных и
- удобство технической реализации.

4.3.3.2 Доступ к данным

Доступ к данным рассматривается в аспекте наличия технологических возможностей для упрощения работы пользователей с терминологическими ресурсами при поиске, считывании и использовании нужных терминологических данных.

Пример — Примером могут служить запросные компьютерные системы или функции, предметные указатели в словарях и т. п.

Реализация таких возможностей требует разнообразных технологий работы с терминологическими ресурсами. Для обеспечения доступа к целевому информационному контенту наиболее важными факторами являются его эффективность и удобство. При осуществлении доступа обычно формируются такие показатели, как доля получаемых точных ответов на запросы и доля не точных, но релевантных ответов. Эти показатели служат мерой эффективности и удобства доступа, а соответствующая им статистика может использоваться для оценки значения данной характеристики полезности.

4.3.3.3 Пользовательский отбор данных

Пользовательский отбор данных — это атрибут, характеризующий наличие возможности для пользователей предопределять содержание выдаваемой по запросу информации, то есть указывать, что они хотят получать в качестве терминологических данных (например, термин, определение, синонимы и др.) из терминологического ресурса. Атрибутами такого отбора являются:

- гибкость отбора категорий данных и
- гибкость структуры выходных данных.

4.3.3.3.1 Гибкий отбор категорий данных

Гибкость отбора категорий данных определяется возможностями для пользователей выбирать те категории, которым должны принадлежать выдаваемые по запросам данные. При этом в действительности возможно задание некоторого подмножества всех категорий данных, охватываемых конкретным терминологическим ресурсом.

Измерениям подлежат два свойства данной характеристики полезности:

- соответствие категории данных нуждам пользователя (см. также 4.3.1.2);
- гибкость допустимых способов отбора нужных категорий данных.

4.3.3.3.2 Гибкая структура выходных данных

Гибкость структуры выходных данных определяет способность терминологического ресурса предоставлять в ответ на запрос выходные данные предписанной структуры. Этот атрибут измерим, когда фактическая структура выходных данных и пользовательская спецификация требуемой структуры данных сопоставимы.

Для поддержки возможности повторного использования данных во многих стандартах используется определенная стандартизованная методология определения требуемой структуры данных в рамках коллекции терминологических данных. В частности, такими стандартами являются ИСО 11179 (*Реестры метаданных*) и ИСО 16642 (*Схема терминологической разметки*), которые играют важную роль, потому что соответствие этим стандартам обеспечивает возможности повторного использования данных и взаимодействия между разными терминологическими ресурсами.

4.3.4 Характеристики полезности способов ввода данных

Характеристики полезности, связанные со способом ввода данных, касаются требований к управлению данными, а также (в большей мере) начальной регистрации терминологических данных. Атрибуты, используемые применительно к этому аспекту оценки ресурсов, должны включать в себя:

- удобный для пользователя формат ввода данных;
- проверку достоверности данных;
- обеспечение совместимости с другими структурами данных.

4.3.4.1 Удобный для пользователя формат ввода данных

Этот атрибут относится к структуре интерфейса, используемого при обработке терминологических данных и их записи в терминологический ресурс.

Для обеспечения удобства использования пользовательского интерфейса необходимо соответствовать требованию читабельности данных. Поэтому при оценке данной характеристики полезности должен приниматься во внимание 4.3.3.1.

4.3.4.2 Проверка достоверности данных

К этой характеристике полезности применимо описание атрибутов, данное в 4.3.2.1.

Прежде чем какие-либо данные определяются как записываемые в терминологический ресурс, должен быть произведен их контроль на достоверность для гарантии того, что входные данные соответствуют общим требованиям терминологического ресурса.

4.3.4.3 Совместимость с другими структурами данных

Совместимость с другими структурами данных нужна для обеспечения возможности обработки данных, подлежащих импортированию из других терминологических ресурсов. В этой связи подлежат анализу два атрибута:

- отображение подлежащих импортированию категорий данных;
- преобразование структуры данных, подлежащих импортированию.

4.3.4.3.1 Отображение подлежащих импортированию категорий данных

Этот атрибут показывает, могут ли категории данных, подлежащие импортированию, быть приняты и обработаны в рамках терминологического ресурса с целью интеграции импортируемых данных. При оценивании этой характеристики полезности следует определить, являются ли импортируемые категории данных стандартизованными или они соответствуют существующим стандартам, например, таким как ИСО 12620 (см. также 4.3.1.2).

4.3.4.3.2 Преобразование структуры данных, подлежащих импортированию

Этот атрибут показывает, может ли структура данных, подлежащих импортированию, быть адаптирована в рамках терминологического ресурса для интеграции импортируемых данных. При оценивании этой характеристики полезности следует определить, соответствует ли структура импортируемых данных известным стандартным моделям, например, таким как определенные в ИСО 16642 (см. также 4.3.1.1).

5 Руководящие принципы оценки и сравнительного анализа

5.1 Общие положения

Существуют два подхода к оценке терминологических ресурсов.

Один из них состоит в том, чтобы, начав с целевого терминологического ресурса, систематически устанавливать характеристики полезности на основе тех атрибутов, которые определены в настоящем стандарте. Далее желаемые характеристики полезности целевого терминологического ресурса должны использоваться как эталоны для сравнения с потенциальными требованиями пользователя.

Второй подход заключается в том, чтобы начать сразу с пользовательских требований, без определения целевого ресурса. Поэтому сначала должны быть уточнены потребности пользователя. Затем подлежат определению характеристики полезности на основе учета конкретных нужд пользователя и положений настоящего стандарта. Установленные характеристики полезности могут далее служить эталонами, с которыми могут сравниваться или по которым могут оцениваться терминологические ресурсы. В рамках второго подхода будет полезно использование проектной информации ИСО 15188.

5.2 Основные этапы работы

5.2.1 Определение целей

Конечной целью проекта по оценке и эталонному анализу должно быть отыскание терминологического ресурса и его эффективное использование. Этот проект должен служить также следующим целям:

- установление уровней практической ценности терминологического ресурса: например, для рейтинговых или сертификационных служб, связанных с терминологическими ресурсами;
- создание или расширение терминологического ресурса для достижения заданного уровня полезности;
- демонстрация пригодности терминологического ресурса для потенциальных пользователей: например, для выбора подходящего терминологического ресурса или для службы поиска ресурсов.

5.2.2 Определение эталонов

В зависимости от конкретной цели проекта на основе использования настоящего стандарта должны быть выбраны в качестве эталонов соответствующие характеристики полезности.

5.2.3 Проверка соответствия эталонам

С эталонами должны сопоставляться потенциальные терминологические ресурсы с целью определения степени их соответствия конкретным целевым требованиям.

5.3 Тестовая оценка

Для проверки соответствия пользовательским требованиям должен быть выбран для использования или обработан небольшой объем выходных данных терминологического ресурса.

Этот терминологический ресурс должен некоторое время использоваться, а затем должен быть оценен на основе реального пользовательского опыта.

Следует иметь в виду, что в заключении по экспериментальной оценке ресурса должны приниматься во внимание также психологические впечатления пользователей и их субъективные ощущения и суждения.

Приложение А
(обязательное)

Характеристики полезности терминологических ресурсов и основные правила их оценки

Характеристики полезности	Основные правила измерения (минимальные требования)
4.3.1 Характеристики полезности, относящиеся к терминологическим данным	
4.3.1.1 Спецификация структуры данных	Явное описание или определение типа документа [DTD (document type definition)] либо схема, которые должны облегчать пользователям понимание или различение единиц терминологической информации. ***Пример — Для удовлетворения потребности в стандартизации терминологии спецификация структур данных должна представлять минимальную структуру статьи, определенную в ИСО 10241, или простейшую модель деривата, основанную на обобщенной спецификации, представленной в ИСО 16642***
4.3.1.2 Охват категорий данных	Категории данных, необходимые пользователям, должны сопоставляться с теми категориями, которые предоставляет терминологический ресурс на основе использования стандартизованных категорий данных или согласованного механизма сравнения. Мерилом может быть процент благоприятных совпадений с пользовательскими категориями
4.3.1.3 Охват предметных областей	Должны быть доступны категории данных областей и подобластей знаний. Целевой терминологический ресурс должен охватывать наиболее весомую для пользователей предметную область. Измеряемым параметром должна быть процентная доля совпадений с элементами пользовательского контрольного списка индикаторов предметных областей либо результат заданной процедуры статистической обработки таких совпадений
4.3.1.4 Охват языков	Должны быть доступны явные описания или категории данных, содержащие идентификаторы языков, которые охватываются ресурсом. Указатели нужных пользователю языков могут проверяться вручную или автоматически путем анализа открытого текста; может также производиться сравнение указателей с использованием стандартизованных кодов из ИСО 639, ИСО 15924 и ИСО 3166 или из других отраслевых стандартов. Результатом анализа может быть таблица ответов «есть» или «нет» применительно к языкам, указанным в пользовательском контрольном списке
4.3.1.5 Соответствие правилам когерентности	Пользователям должны быть предельно ясны правила, по которым функционирует терминологический ресурс; для этого в нем должны присутствовать описательные инструкции или специальные категории данных. Для оценки этого атрибута должны приниматься во внимание два аспекта: требуется ли пользователю когерентность и если да, то степень соответствия терминологического ресурса этому пользовательскому требованию
4.3.1.6 Использование контролируемых внешних данных	Этот атрибут особенно подходит для электронных терминологических ресурсов. В рамках ресурса должны существовать специальные категории данных, позволяющие устанавливать соответствующие указатели и ссылки. Для оценки значений данного атрибута должны быть встроены надлежащие механизмы автоматической проверки таких указателей и ссылок. Мерилом использования контролируемых внешних данных может служить сохранение целостности конкретных указателей и ссылок

Продолжение

Характеристики полезности	Основные правила измерения (минимальные требования)
4.3.1.7 Использование авторитетных источников	Оценке уровня авторитетности используемых источников либо установлению факта наличия или отсутствия такого использования должен помогать качественный анализ указанных либо заранее определенных источников на основе анализа контента в виде значений категорий библиографических данных, содержащихся в терминологическом ресурсе
4.3.1.8 Указание обладателя прав интеллектуальной собственности	Если информация защищена авторским правом, то для ее использования обычно требуется ознакомление с директивным описанием юридического характера или переход по ссылке к сведениям о правообладателе. Оценка этого атрибута проясняет в качественной форме права сторонних лиц на использование защищенных терминологических данных терминологического ресурса
4.3.1.9 Симметрия представления коллекции терминологических данных	Этот атрибут особенно важен для терминологических ресурсов, которые состоят из разделов, относящихся к разным языкам. Для оценки этого атрибута необходимо принимать в расчет степень сбалансированности контента по ключевым категориям данных, образующим информационную структуру ресурса. Результатом предписанных статистических вычислений может быть определение пустых и наполненных категорий. Такие расчеты применительно к нужным пользователю категориям данных могут служить основой для построения метрики оценок. Для выполнения автоматических измерений необходимо иметь категории данных, которые определяют состояние наполнения ресурса и состояние контрольной проверки
4.3.1.10 Объем коллекции терминологических данных	Терминологический ресурс должен содержать в себе довольно большой объем информации. Метрикой является «величина» ресурса, то есть объем информационного контента для требуемых пользователем категорий данных
4.3.2 Характеристики полезности, связанные с механизмом управления данными	
4.3.2.1 Проверка достоверности данных	Пользователям должны быть ясны реализованные в ресурсе правила контроля достоверности данных. Используя установленные правила, пользователь может проверять качество предоставляемых ему терминологических данных. В случае электронных ресурсов может оказаться доступной и автоматическая проверка эффективности действующих правил
4.3.2.1.1 Полнота	Здесь оценивается наличие или отсутствие в терминологическом ресурсе надлежащих механизмов контроля и управления. При оценке значений данного атрибута может также использоваться механизм, описанный в 4.3.1.9 (данного приложения)
4.3.2.1.2 Лингвистическая корректность	Измерение этого атрибута показывает, имеются или не имеются в терминологическом ресурсе необходимые механизмы контроля и управления
4.3.2.1.3 Соответствие условиям представления	Терминологические данные должны соответствовать соглашению о способах представления данных, определенному в том терминологическом ресурсе, где они записаны и хранятся. Если такое соглашение отсутствует, то предоставляемые терминологическим ресурсом терминологические данные должны соответствовать требованиям пользователя. Анализ значений целевых категорий данных для получения ответа на вопрос, удовлетворяются или не удовлетворяются заданные пользователем стандарты представления данных, может применяться ручная процедура или компьютерная программа. Мерилом может быть качественное описание требований или количественный расчет на основе известных требований пользователя.
4.3.2.2 Проверка на избыточность	Для реализации этой проверки должно быть декларировано условие обнаружения избыточности данных

Продолжение

Характеристики полезности	Основные правила измерения (минимальные требования)
Контроль уровня дублирования понятий	Когда пользователем определен диапазон нужных предметных областей, существует возможность количественной оценки данного атрибута (в ручном режиме или автоматически, с помощью прикладных программ) путем сравнения терминов или определений. В конечном итоге определяется измеримый статистический показатель дублирования понятий
4.3.2.3 Регулярное сопровождение данных (поддержание в актуальном состоянии)	Качественной оценкой этого атрибута является заключение о наличии или отсутствии механизмов такого сопровождения. Проверка значений категорий данных по датам может дать еще одно измерение. ***Пример — Список регулярных изданий, выпускаемых типографским способом, показывает, что данные этого ресурса постоянно просматриваются и обновляются***
4.3.2.4 Регулярное сопровождение метаданных	Спецификация метаданных должна быть все время открытой для пользователей в целях приспособления их к пользовательским нуждам. Качественной метрикой этого атрибута является наличие или отсутствие таких механизмов
4.3.3 Характеристики полезности, относящиеся к выводу данных	
4.3.3.1 Читабельность выходных данных	Выходные данные должны обеспечивать пользователям возможность чтения или обработки информации. Компонентами метрики могут быть пользовательские требования в отношении механизма доступа к выходным данным, соответствие пользовательским стандартам или «лучшие достижения сложившейся практики» в области терминологической работы. ***Пример — ИСО 10241:1992 определяет структуру терминологических статей для терминологического стандарта ИСО***
4.3.3.2 Доступ к данным	Должен существовать эффективный способ доступа пользователей к данным целевого терминологического ресурса. ***Пример — Для обеспечения возможности работы с терминологическими ресурсами китайского языка, изданными типографским способом, обязательным условием является наличие в них указателя терминов.*** Метрикой может служить статистический показатель доли получаемых точных и релевантных ответов в выборке запросов. Однако качественная оценка степени точности и релевантности должна оставаться прерогативой пользователя. ***Пример — Пользователь может подготовить заранее ряд запросов с заведомо известными элементами ответов. На этой основе пользователь может обращаться с запросами к целевому терминологическому ресурсу в ручном режиме либо с помощью компьютерной программы, если это возможно, и затем дать качественную или даже количественную оценку***
4.3.3.3 Пользовательский отбор данных	Пользователь должен иметь возможность создавать подмножество собственных категорий данных.
4.3.3.3.1 Гибкий отбор категорий данных	Отбор разрешенных категорий выходных данных терминологического ресурса должен производиться в соответствии с информационными нуждами пользователя. При оценке значений данного атрибута следует использовать также механизм, описанный в 4.3.1.2.
4.3.3.3.2 Гибкая структура выходных данных	Для обеспечения возможности сопоставления структуры выходных данных с требованиями пользователя должно существовать явное описание структуры или определение типа документа (DTD) или соответствующая схема. Для оценки этого атрибута следует учитывать 4.3.1.1.

Окончание

Характеристики полезности	Основные правила измерения (минимальные требования)
4.3.4 Характеристики полезности способов ввода данных	
4.3.4.1 Удобный для пользователя формат ввода данных	Важнейшим аспектом оценки должна считаться читабельность (см. 4.3.3.1)
4.3.4.2 Проверка достоверности данных	См. 4.3.2.1
4.3.4.3 Совместимость с другими структурами данных	Структура целевого терминологического ресурса должна быть совместима с данными, имеющими иную структуру
4.3.4.3.1 Отображение подлежащих импортированию категорий данных	Категории данных целевого терминологического ресурса должны соответствовать категориям данных, выбранным пользователем. (См. 4.3.1.2 и 4.3.3.3.1)
4.3.4.3.2 Преобразование структуры данных, подлежащих импортированию	Этот атрибут особенно важен в применении к управляемым компьютеризованным терминологическим ресурсам. Для того чтобы терминологический ресурс обладал такими свойствами полезности, пользователям должно быть разрешено определение собственного множества структурных элементов. (См. 4.3.1.1 и 4.3.3.3.2). Элементами метрики могут быть значения «совместим» и «несовместим» по отношению к явному пользовательскому описанию, пользовательскому DTD или пользовательской схеме

Приложение В
(справочное)

Типовое использование характеристик полезности как эталонов

В.1 Общий обзор

При использовании того или иного терминологического ресурса пользователей прежде всего волнуют вопросы адекватности контента, структурной совместимости и качества формализованного представления терминологических данных. В связи с этим ниже предлагаются рекомендуемые общие принципы оценки и сравнительного анализа терминологических ресурсов.

В.2 Стандартная оценка и сравнительный анализ терминологических ресурсов

В.2.1 Соответствие потребностей качеству контента

В.2.1.1 Соответствие реальным потребностям

Факт наличия или отсутствия в терминологическом ресурсе действительно нужных пользователю терминологических данных может быть установлен посредством оценки следующих критериев контроля:

- охват категорий данных (4.3.1.2);
- охват предметных областей (4.3.1.3);
- охват языков (4.3.1.4);
- симметрия представления совокупности терминологических данных (4.3.1.9);
- объем совокупности терминологических данных (4.3.1.10).

Предметные области, языки и категории данных, покрываемые терминологическим ресурсом, оцениваются для того, чтобы максимизировать степень их соответствия потребностям пользователя. Чем больше степень соответствия, тем выше оказывается результат оценивания ресурса.

Если применительно к трем первым критериям не устанавливается никаких специфических требований, то чем шире охват у терминологического ресурса, тем выше вероятность, что данный терминологический ресурс будет соответствовать терминологическому контенту, необходимому пользователям.

В.2.1.2 Качество контента

В контексте данного приложения слово «качество» относится к качеству терминологических данных, содержащихся в терминологическом ресурсе. При этом рассматриваются два аспекта: 1) уровень качества в ракурсах лингвистики и языка и 2) уровень качества в контексте специальных знаний.

Критериями оценки терминологических данных могут служить характеристики полезности, относящиеся к качеству терминографической работы. Соответствие этим критериям должно косвенно отражать в какой-то степени качество в разрезе специальных знаний и охвата языков. К числу таких критериев могут относиться:

- соответствие правилам когерентности (4.3.1.5);
- использование контролируемых внешних данных (4.3.1.6);
- использование авторитетных источников (4.3.1.7);
- лингвистическая корректность (4.3.2.1.2);
- соответствие условиям представления (4.3.2.1.3);
- проверка на избыточность (4.3.2.2).

Эти критерии могут применяться к выходной информации терминологических ресурсов для оценки их качества.

Если терминологический ресурс имеет механизмы контроля вышеуказанных аспектов, например механизм контроля достоверности при вводе информации или управлении данными, то есть все основания надеяться на получение высококачественных данных.

В дополнение к перечисленным критериям может использоваться выборка данных заведомо высокого качества — как эталон, с которым должны сравниваться выходные данные терминологического ресурса. Подобное сравнение — это тоже один из методов оценки качества ресурсов.

В.2.2 Оценка уровня терминологических знаний

Данный атрибут касается, в основном, характеристик полезности выходной информации. Оценочными критериями для данного аспекта полезности могут служить следующие характеристики полезности:

- указатель обладателя (принадлежности) прав интеллектуальной собственности (4.3.1.8);
- читабельность выходных данных (4.3.3.1);
- доступ к данным (4.3.3.2),
- пользовательский отбор данных (4.3.3.3).

Пример 1 — Система управления базами данных (СУБД) с расширенными возможностями поиска терминологической информации позволит пользователям запрашивать гораздо больший объем полезных данных.

Пример 2 — Высококачественный словарь будет иметь, помимо многочисленных указателей, еще и четко структурированное расположение статей, которые легко читать, и в котором очень понятно выделяются разделы основного содержания. А терминологические и прочие стандарты ИСО имеют четко выделенные разделы терминологии, соответственно спецификациям ИСО 10241.

Пользовательский отбор данных об интерфейсах будет возможен только применительно к электронным терминологическим ресурсам.

Те, кто имеет намерение использовать данные терминологического ресурса в коммерческих целях, должны иметь четкое представление о правообладателе терминологических данных, которыми они планируют воспользоваться.

В.2.3 Обмен терминологическими данными

Обмен терминологическими данными происходит между людьми, организациями или информационными системами. При использовании терминологических данных для целей информационного обмена могут рассматриваться следующие атрибуты:

- сведения о владельце прав интеллектуальной собственности (4.3.1.8),
- возможности пользовательского отбора данных (4.3.3.3);
- возможность проверки достоверности данных (4.3.4.2);
- сведения о совместимости используемого ресурса с другими структурами данных (4.3.4.3).

Для обеспечения возможности обмена терминологическими данными главенствующую роль играет разрешение проблемы авторского права.

Требуемые категории данных для разных видов информационного обмена могут различаться. Задача организации обмена данными упрощается, если возможен отбор нужных категорий данных и процедура отбора обладает достаточной гибкостью.

Форматы данных, структурированных в соответствии с метамоделью, определены в ИСО 16642; категории данных, установленные в ИСО 12620, и разметка XML, представленная в ИСО 30042, способны взаимодействовать друг с другом, то есть возможно преобразование данных из одного формата в другой без потери информации. Согласно положению 5.5 ИСО 16642:2003, «при следовании описанному здесь интегрированному подходу велика вероятность, что результирующий формат будет легко адаптироваться к изменяющимся условиям и будет совместим с другими структурами баз данных и другими форматами».

В.2.4 Обеспечение процессов терминологической работы

Терминологическая работа может рассматриваться в двух разрезах:

a) Терминологическая работа обычно поддерживается терминологическим ресурсом, каковым является автоматизированная система обработки информации. (В прежние времена это могли быть карточные файлы или рукописи в скоросшивателях.)

b) В ходе работы с терминологией часто порождаются новые терминологические ресурсы для других пользователей.

Когда терминологический ресурс предназначается для поддержки крупномасштабной терминологической работы, к атрибутам, приведенным в В.2.1—В.2.3, могут быть добавлены:

- критерии полезности способов ввода данных:
  - удобный для пользователя формат ввода данных (4.3.4.1);
  - наличие проверки достоверности данных (4.3.4.2);
  - совместимость с другими структурами данных (4.3.4.3),
- критерии, относящиеся к качеству контента; они могут быть ориентированы на установление факта наличия или отсутствия контроля следующих характеристик полезности:
  - проверки достоверности данных (4.3.2.1);
  - проверки на избыточность (4.3.2.2).

Критерии, ориентированные на оценку конечного продукта терминологической работы, основываются на анализе конкретного целевого назначения создаваемого продукта.

В.2.5 Семантическая паутина

Применительно к технологии семантической паутины терминологические данные используются для «понимания» сведений о знаниях; необходимая для этого оценка формируется автоматически. При этом могут учитываться следующие характеристики полезности (определяющие познавательную ценность и доступность данных):

- спецификация структуры данных (4.3.1.1);
- охват категорий данных (4.3.1.2);
- охват предметных областей (4.3.1.3);

- охват языков (4.3.1.4);
- проверка на избыточность (4.3.2.2);
- регулярное сопровождение данных (4.3.2.3);
- регулярное сопровождение метаданных (4.3.2.4);
- доступ к данным (4.3.3.2).

Важную роль играет стандартизованное представление данных, относящихся к этим атрибутам.

В.2.6 Слияние одновременно используемых распределенных ресурсов данных

Все чаще услугами распределенных терминологических ресурсов начинают пользоваться оперативно доступные системы ассоциативного обслуживания. Очень важно обеспечить успешное объединение таких терминологических ресурсов — прежде всего потому, что данные различных ресурсов могут совпадать (дублироваться). В этом объединении важную роль играют следующие характеристики полезности:

- спецификация структуры данных (4.3.1.1),
- охват категорий данных (4.3.1.2);
- охват предметных областей (4.3.1.3);
- охват языков (4.3.1.4);
- указание обладателя прав интеллектуальной собственности (4.3.1.8);
- регулярное сопровождение данных (4.3.2.3);
- регулярное сопровождение метаданных (4.3.2.4);
- доступ к данным (4.3.3.2);
- пользовательский отбор данных (4.3.3.3).

Использование стандартных подходов, в рамках которых принимаются во внимание взаимосвязанные характеристики полезности, будет гарантировать эффективное взаимодействие между различными ресурсами.

Приложение ДА
(справочное)

Сведения о соответствии ссылочных международных стандартов национальным стандартам Российской Федерации

Таблица ДА.1

Обозначение ссылочного международного стандарта	Степень соответствия	Обозначение и наименование соответствующего национального стандарта
ИСО 12616:2002	—	*
ИСО 12620:2009	IDT	ГОСТ Р ИСО 12620—2012 «Терминология, другие языковые ресурсы и ресурсы содержания. Спецификация категорий данных и ведение реестра категорий данных для языковых ресурсов»
ИСО 16642:2003	—	*
ИСО 3166 (все части)	—	*
ИСО/МЭК 11179 (все части)	IDT	ГОСТ Р ИСО/МЭК 11179 «Информационная технология. Регистры метаданных (РМД)»
ИСО/МЭК ТО 9126 (все части)	—	*
ИСО 9241-11:1998	IDT	ГОСТ Р ИСО 9241-11—2010 «Эргономические требования к проведению офисных работ с использованием видеодисплейных терминалов (VDT). Часть 11. Руководство по обеспечению пригодности использования»
ИСО 639 (все части)	—	*
ИСО 15924:2004	—	*
ИСО 10241 (все части)	—	*
ИСО 860:2007	—	*
ИСО 15188:2001	IDT	ГОСТ Р ИСО 15188—2012 «Принципы управления проектами стандартизации терминологии»

* Соответствующий национальный стандарт отсутствует. До его утверждения рекомендуется использовать перевод на русский язык данного международного стандарта. Перевод данного международного стандарта находится в Федеральном информационном фонде технических регламентов и стандартов.

Примечание — В настоящей таблице использовано следующее условное обозначение степени соответствия стандартов:

- IDT — идентичные стандарты.

Библиография

[1] ИСО 639 (все части) — Коды для представления названий языков
(ISO 639 (all parts)) — (Codes for the representation of names of languages)

[2] ИСО 704:2009 — Терминологическая деятельность. Принципы и методы
(ISO 704:2009) — (Terminology work — Principles and methods)

[3] ИСО 860:2007 — Терминологическая работа. Гармонизация понятий и терминов
(ISO 860:2007) — (Terminology work — Harmonization of concepts and terms)

[4] ИСО 1087-1:2000 — Терминологическая работа. Словарь. Часть 1. Теория и применение
(ISO 1087-1:2000) — (Terminology work. Vocabulary. Part 1. Theory and application)

[5] ИСО 3166 (все части) — Коды для представления названий стран и единиц их административно-территориального деления
(ISO 3166 (all parts)) — (Codes for the representation of names of countries and their subdivisions)

[6] ИСО 9241-11:1998 — Эргономические требования, связанные с использованием видеотерминалов для учрежденческих работ. Часть 11. Руководство по определению и измерению используемости
(ISO 9241-11:1998) — (Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability)

[7] ИСО 10241 (все части) — Терминологические статьи в стандартах
(ISO 10241 (all parts)) — (Terminological entries in standards)

[8] ИСО/МЭК 11179 (все части) — Информационные технологии. Реестры метаданных (MDR)
(ISO/IEC 11179 (all parts)) — (Information technology — Metadata registries (MDR))

[9] ИСО 12616:2002 — Терминография, ориентированная на перевод
(ISO 12616:2002) — (Translation-oriented terminography)

[10] ИСО 12620:2009 — Терминология, другие языковые ресурсы и ресурсы содержания. Спецификация 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 resources)

[11] ИСО 15188:2001 — Принципы управления проектами стандартизации терминологии
(ISO 15188:2001) — (Project management guidelines for terminology standartisation)

[12] ИСО 15924:2004 — Информация и документация. Коды для представления названий шрифтов
(ISO 15924:2004) — (Information and documentation — Codes for the representation of names of scripts)

[13] ИСО 16642:2003 — Применение компьютера в терминологических целях. Структура терминологической разметки
(ISO 16642:2003) — (Computer applications in terminology — Terminological markup framework)

[14] ИСО 30042:2008 — Системы для управления терминологией, знаниями и содержанием. Обмен базами данных (TermBase eXchange (TBX))
(ISO 30042:2008) — (Systems to manage terminology, knowledge and content — TermBase eXchange (TBX))

[15] Директивы ИСО/МЭК, часть 2 — Правила построения и изложения международных стандартов

[16] ИСО/МЭК 2382-17:1999 — Информационные технологии. Словарь. Часть 17. Базы данных
(ISO/IEC 2382-17:1999) — (Information technology — Vocabulary — Prt 17. Databases)

[17] ИСО/МЭК 9126 (все части) — Программирование. Качество продукта
(ISO/IEC 9126 (all parts)) — (Software engineering — Product quality)

[18] BCP 47, IETF Industry Standard

УДК 001.4:006.354 ОКС 01.020 Т58

Ключевые слова: терминология, терминологический ресурс, принципы и требования к терминологическим системам, характеристики полезности

Редактор *Т.С. Никифорова*
Технический редактор *В.Н. Прусакова*
Корректор *Е.Д. Дульнева*
Компьютерная верстка *А.Н. Золотаревой*

Сдано в набор 18.04.2015. Подписано в печать 30.04.2015. Формат 60 × 84 1/8. Гарнитура Ариал.
Усл. печ. л. 2,79. Уч.-изд. л. 2,30. Тираж 36 экз. Зак. 1817.

Издано и отпечатано во ФГУП «СТАНДАРТИНФОРМ», 123995 Москва, Гранатный пер., 4.
www.gostinfo.ru gost.e-ecolog.ru

gost.e-ecolog.ru