
**Управление языковыми ресурсами.
Система синтаксического
аннотирования (SynAF).
Часть 1. Синтаксическая модель**

*Language resource management. — Syntactic annotation framework
(SynAF) —
Part 1: Syntactic model*

get full document from standards.iteh.ai

Ответственность за подготовку русской версии несёт GOST R
(Российская Федерация) в соответствии со статьёй 18.1 Устава ISO



Ссылочный номер
ISO 24615-1:2014

© ISO 2014

Sample Document

get full document from standards.iteh.ai



ДОКУМЕНТ ЗАЩИЩЕН АВТОРСКИМ ПРАВОМ

© ISO 2014

Все права сохраняются. Если не указано иное, никакую часть настоящей публикации нельзя копировать или использовать в какой-либо форме или каким-либо электронным или механическим способом, включая фотокопии и микрофильмы, без предварительного письменного согласия ISO по соответствующему адресу, указанному ниже, или комитета-члена ISO в стране заявителя.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Опубликовано в Швейцарии

Содержание

Страница

Предисловие	iv
Введение	v
1 Область применения	1
2 Нормативные ссылки	1
3 Термины и определения	1
4 Мета модель SynAF	4
4.1 Вводные замечания	4
4.2 Описание метамодели SynAF	5
Приложение А (нормативное) Категории данных для метамодели SynAF	7
Приложение В (информативное) Связь с системой лингвистического аннотирования	16
Библиография	18

Sample Document

get full document from standards.iteh.ai

Предисловие

Международная организация по стандартизации (ISO) является всемирной федерацией национальных организаций по стандартизации (комитетов-членов ISO). Разработка международных стандартов обычно осуществляется техническими комитетами ISO. Каждый комитет-член, заинтересованный в деятельности, для которой был создан технический комитет, имеет право быть представленным в этом комитете. Международные правительственные и неправительственные организации, имеющие связь с ISO, также принимают участие в работе. ISO работает в тесном сотрудничестве с Международной электротехнической комиссией (IEC) по всем вопросам стандартизации в области электротехники.

Процедуры разработки, использованные при создании настоящего документа, и надлежащие операции его дальнейшего сопровождения описаны в Части 1 Директив ISO/IEC, в которых особенно важны критерии утверждения и приёмы различных документов ISO. Проект настоящего документа был подготовлен в полном соответствии с правилами, приведёнными в Части 2 Директив ISO/IEC (см. www.iso.org/directives).

Принимается во внимание тот факт, что некоторые из элементов данного документа могут быть объектом патентных прав, однако ISO не принимает на себя никаких обязательств по определению отдельных или всех таких прав. Подробные сведения о патентных правах, выявляемых в процессе разработки того или иного документа, подлежат включению в его введение и/или в перечень полученных ISO деклараций (см. www.iso.org/patents).

Любая торговая марка, фигурирующая в настоящем документе, является информацией, представляемой исключительно для удобства пользователей стандарта, и не может считаться признаком выражаемого предпочтения.

Разъяснение специфических терминов и формулировок ISO, касающихся оценки соответствия конкретным нормативным документам, равно как и приверженности ISO принципам Всемирной торговой организации (ВТО) в части устранения технических барьеров в торговле, можно найти на сайте ISO по унифицированному указателю информационного ресурса (URL) [Foreword — Supplementary information](#).

Настоящий документ был подготовлен подкомитетом SC 4 “Управление языковыми ресурсами” Технического комитета ISO/TC 37 “Терминология и другие языковые и информационные ресурсы”.

Настоящее первое издание ISO 24615-1 отменяет и заменяет собой стандарт ISO 24615:2010, претерпевший минимальные изменения.

ISO 24615 (во всех его частях) предназначен для использования в сочетании со стандартами ISO 24612 “Управление языковыми ресурсами. Система лингвистического аннотирования (LAF)”, ISO 24613:2008, “Управление языковыми ресурсами. Схема лексической разметки (LMF)” и ISO 24611 “Управление языковыми ресурсами. Система морфосинтаксического аннотирования (MAF)”.

ISO 24615 состоит из следующих частей, объединённых общим заголовком “Управление языковыми ресурсами. Система синтаксического аннотирования (SynAF)”:

— Часть 1. Синтаксическая модель

и находящаяся в стадии разработки

— Часть 2. Индексирование XML-документов (<Tiger2/>).

Введение

В основе стандартов серии ISO 24615 лежат многочисленные проекты и мероприятия последнего десятилетия, предшествовавшие стандартизации [9] и обеспечившие создание различных моделей и форматов представления синтаксической информации в виде механизмов синтаксического анализа или аннотирования языковых ресурсов (корпусов с синтаксической разметкой). В течение целого ряда лет стандартом де-факто для процедур синтаксического аннотирования служил инициативный проект Penn Treebank (синтаксически аннотированный текстовый корпус Пенсильванского университета). Однако в более поздних работах — таких как инициативный проект Negra/Tiger в Германии (см. <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/>) и проект ISST в Италии [18] — были продемонстрированы более связные, более согласованные и более понятные системы синтаксического анализа, способные в равной степени учитывать как отношения иерархической соподчинённости компонентов, так и явление зависимости в синтаксическом аннотировании.

Особенно плодотворным в этом смысле оказался проект электронных информационных ресурсов “LIRICS”, который позволил объединить усилия группы экспертов, инициировавшей разработку проекта Международного стандарта ISO 24615, касающегося системы синтаксического аннотирования (SynAF). Эта группа в процессе разработки системы SynAF установила тот факт, что в действительности в рамках всех перечисленных выше инициативных проектов используется общая модель данных, которая создаёт хорошую основу для построения метамодели системы синтаксического аннотирования [см. результаты исследований, представленные в информационном бюллетене Deliverable D.3.1 по проекту Европейского Союза LIRICS; эта публикация касается оценки инициативных проектов систем морфосинтаксического и синтаксического аннотирования (“Evaluation of initiatives for morpho-syntactic and syntactic annotation”) и размещена по адресу http://lirics.loria.fr/doc/pub/Del3_1_V2.pdf].

В настоящей части ISO 24615 предлагается метамодель синтаксического аннотирования совместно с перечнем релевантных категорий данных для синтаксического аннотирования. Соответствующие категории данных доступны на сервере ISOCat (<http://www.isocat.org/>) в синтаксическом профиле (как он определён в ISO 12620:2009).

Sample Document

get full document from standards.iteh.ai

Управление языковыми ресурсами. Система синтаксического аннотирования (SynAF).

Часть 1.

Синтаксическая модель

1 Область применения

Первая часть ISO 24615 содержит описание системы синтаксического аннотирования (SynAF), которая является высокоуровневой моделью для представления синтаксической аннотации лингвистической информации в целях обеспечения надлежащего взаимодействия между языковыми ресурсами или компонентами процесса обработки языковых данных. Эта часть ISO 24615 служит дополнением стандарта ISO 24611 (относящегося к системе морфосинтаксического аннотирования) и тесно связана с ним, поскольку предоставляет метамодель для формирования синтаксических представлений и эталонные категории данных для отображения местоположения и зависимости информационных объектов в рамках сложных предложений или других сопоставимых высказываний и сегментов.

2 Нормативные ссылки

Перечисленные ниже ссылочные документы обязательны для применения данного документа. В случае датированных ссылок действующим является только указанное издание. Применительно к недатированным ссылочным документам применяются их самые последние издания (включая все последующие изменения):

ISO 1087-1:2000, *Терминологическая работа. Словарь. Часть 1. Теория и применение*

ISO 12620:2009, *Терминология, другие языковые ресурсы и ресурсы содержания. Спецификация категорий данных и ведение реестра категорий данных для языковых ресурсов*

ISO 24611:2012, *Управление языковыми ресурсами. Система морфосинтаксического аннотирования*

3 Термины и определения

Для целей данного документа используются термины и определения из стандартов ISO 1087-1:2000, ISO 12620:2009, ISO 24611:2012, а также терминология, приведённая ниже.

3.1

обстоятельственное слово, обстоятельство, адъюнкт
adjunct

второстепенный элемент, ассоциируемый с глаголом, в отличие от *синтаксических аргументов* (3.19)

Примечание 1 к словарной статье: Обстоятельными словами в предложении могут быть наречия.

3.2

фрагмент
chunk

нерекурсивная *составляющая* (3.4)

3.3

клауза, элементарное предложение
clause

группа *фраз* (3.14), обычно содержащая некоторое высказывание