



ЭНЦИКЛОПЕДИЧЕСКАЯ ИНФОРМАЦИЯ О СЛОВЕ И ЛЕКСИЧЕСКИЕ КЛАССЫ

Семенова Софья Юльевна

О СТРУКТУРЕ СЛОВАРЯ РУСЛАН. ПРИМЕР СТАТЬИ

ЗГЛ = ПРЕДВЕСТНИК

КАТ = ЭТК.ОБ, АСПЕКТ

ГХ = СУЩ : ИГ

АНГ = forerunner; prognosticator; signifier

ПО = общ

ВЕС = 4 (А2)

СХ = ФУНКЦИЯ & S1 (предвещать) & СТАДИАЛЬНОСТЬ

ВАЛ = ИДЕНТИФИКАТОР, А1, С

СОДЕРЖАНИЕ, А2, С

ГХ1 = прим_опр : ИГ

СХ1 = АБСТРАКТНОЕ; СИТУАЦИЯ; ЯВЛЕНИЕ;

ОДУШЕВЛЕННОЕ

ГХ2 = κ_доп : Р

СХ2 = СИТУАЦИЯ; ЯВЛЕНИЕ; ПРОЦЕСС; СТАДИАЛЬНОСТЬ;

ВРЕМЯ; Antibon

АСЦ = предвосхищать; предвидеть; буревестник

ИЛЛ = «Оттепель» рубежа 1950-60-х гг. была предвестником

перестройки;

предвестники землетрясений;

предвестники зимы



Энциклопедические функции, наличествующие в словаре РУСЛАН

Anti – антоним: Anti (горячий) = холодный
Anti (горячий) = прохладный
Anti (гуманитарный) =
естественный
Anti (хирург) = терапевт

Cap – главный/выделенный элемент:

Cap (храм) = алтарь

Cap (университет) = ректор

Param – параметр:

Param (электроприбор) = мощность

Param (город) = количество жителей

Param (краска) = цвет



Mesur - единица изменения:

Mesur (сопротивление) = ом

Sing – единичный элемент, квант:

Sing (песок) = песчинка

БОЛЬШОЙ_РАЗМЕР:

БОЛЬШОЙ_РАЗМЕР (камень) = глыба

ПРИНАДЛЕЖНОСТЬ:

ПРИНАДЛЕЖНОСТЬ (хоккеист) = клюшка

ПРИНАДЛЕЖНОСТЬ (аллея) = дерево

СОСТАВНАЯ_ЧАСТЬ: СОСТАВНАЯ_ЧАСТЬ (лестница) = ступень

СОСТАВНАЯ_ЧАСТЬ (биология) = зоология

ИСХОДНАЯ_ТОЧКА: ИСХОДНАЯ_ТОЧКА (детство) = рождение

КОНЕЧНАЯ_ТОЧКА: КОНЕЧНАЯ_ТОЧКА (детство) = юность

ВЫШЕ_ГЕО: ВЫШЕ_ГЕО (залив) = море

(ВЫШЕ_ГЕО (Сибирь) = Россия)

НИЖЕ_ГЕО: НИЖЕ_ГЕО (река) = приток

(НИЖЕ_ГЕО (Российская Федерация) = Московская область)

ЖИТЕЛЬ: ЖИТЕЛЬ (село) = сельчанин

(ЖИТЕЛЬ (Париж) = парижанин)



НЕКОТОРЫЕ ДОПОЛНИТЕЛЬНЫЕ ЭНЦИКЛОПЕДИЧЕСКИЕ ФУНКЦИИ:

Subst – материал изготовления/ вещество:

Subst (сабля) = сталь

Content – содержание (физическое / информационное):

Content (морковь) = каротин

Content (благосостояние) = {обеспеченность; материальные блага; духовные блага; предметы; услуги; условия}

Usage – сфера применения:

Usage (барий) = рентгенодиагностика

Manif – проявление: Manif (воспаление) = отек

Loc – типовое местонахождение: Loc (рыба) = водоем

Loc (медведь) = берлога



Arrar - измерительный прибор:

Arrar (температура) = термометр

Suppl – дополнение: Suppl (скрипка) = смычок

Suppl (чашка) = блюдце

Custom - обычай: Custom (именины) = подарок

РЗВДН - разновидность:

РЗВДН (сварка) = дуговая; плазменная; лазерная;
газовая...

Выше по отраслевому «аспекту»:

ВЫШЕ_ЛИНГВ (германский) = индоевропейский

ВЫШЕ_БИОЛ (укроп) = зонтичные



Что считать энциклопедической информацией?

Энциклопедической (внелингвистической, онтологической) будем считать такую информацию о реалии, а также о комплексе «слово-понятие-реалия», которая:



1) выходит за пределы толкования, т.е. минимальной информации, объясняющей значение

не энциклопедично: велосипед приводится в движение ногами/педалями; см. МАС и Словарь С.И. Ожегова

2) выходит за рамки синтаксического и семантического (категориального) описания валентностей

не энциклопедично:

благодарят – > за ДЕЙСТВИЕ / ОТНОШЕНИЕ, ХОРОШЕЕ;
рабочий работает в ОРГАНИЗАЦИИ,
ср. рабочий – > сборочного цеха

3) выходит за пределы тезаурусных связей ВЫШЕ и СИНОНИМ

не энциклопедично: вода это жидкость;

а лингвистика также называется языкознанием

4) выходит за пределы устойчивых лексических связей, выражаемых лексическими функциями

не энциклопедично: сильный дождь часто называют проливным



ЭНЦИКЛОПЕДИЧНО:

- у велосипеда привод от двух педалей осуществляется через цепную передачу (Энци. словарь 1987 г.);
- в знак благодарности иногда преподносят цветы (культурологическая информация);
- вода – компонент большинства технологических процессов (Энци. словарь 1987 и 2001 гг.);
- в древности серьезный вклад в языкознание внес Аристотель (Энци. словарь 1987 и 2001 гг.);
- диаметр дождевых капель – от 6-7 до 0,5 мм (Энци. словарь 1987 и 2001 гг.).



Энциклопедическая функция «ПАРАМЕТР» - ФОРМАЛИЗМ ДЛЯ ИСЧИСЛЕНИЯ СВОЙСТВ РЕАЛИИ

Некоторые типы декларативной информации:

- X есть Z
 $\Rightarrow Z$ - компонент толкования или тезауруса;
- X подразделяется на W_1, W_2, \dots, W_m
 $\Rightarrow W_1$ есть X, \dots, W_m есть X
- X обладает свойствами Y_1, Y_2, \dots, Y_n
 $\Rightarrow Y_1$ есть компонент толкования
(X есть Z , обладающий свойством Y_1),
а Y_2, \dots, Y_n – энциклопедические данные.



УКАЗАНИЕ ЗНАЧЕНИЙ ПАРАМЕТРОВ В ЭНЦИКЛОПЕДИЧЕСКОЙ ЗОНЕ СЛОВАРЯ

- точные количественные значения:

Энц (сентябрь) = {Param (продолжительность);
продолжительность (сентябрь) = {«точно», 30, день}}

- приближенные количественные значения:

Энц (свет) = {Param (скорость); скорость (свет) =
{«приблизительно», $3 \cdot 10^8$, метр/секунда}}

- интервальные количественные значения:

Энц (дождь) = {Sing(капля);
Param (Sing (дождь)) = диаметр; диаметр (Sing (дождь)) =
«от-до», 0.5, 7, миллиметр}}

- неколичественные значения:

Энц (коттедж) = {Param (материал); материал (коттедж) =
брус; доска; бревно; кирпич, пеноблок,...}



НЕКОТОРЫЕ ЛЕКСИЧЕСКИЕ КЛАССЫ И ИХ ПАРАМЕТРЫ

- резервуары (*канистра, сосуд* и т.п.) – емкость, форма, материал,...
- внутренние помещения (*комната, коридор* и т.п.) – площадь, форма, высота потолка, освещенность, расположение окон, отделка стен,...
- декоративные предметы (*люстра, браслет* и т.п.) – размер, цвет, форма, материал, ценность,...
- одушевленные существа (*ребенок, старик* и т.п.) – возраст, состояние здоровья,...



- строения (*здание, особняк* и т.п.) высота, этажность, конструктивное решение, форма, цвет, материал, назначение,...
- предметы одежды и обуви (*блузка, ботинки* и т.п.) «адресат» (мужское/женское/детское), размер, фасон, цвет, волокно, производитель,...
- культурные растения (*яблоня, черешня* и т.п.) ареал культивирования, сроки созревания плодов, содержание полезных веществ в плодах, зимостойкость, урожайность...
- строительные материалы (*кирпич, цемент* и т.п.) прочность, состав, сфера применения, вид (марка)...
- информационные продукты (*книга, статья* и т.п.) объем



РОЛЬ ЭНЦИКЛОПЕДИЧЕСКОЙ ИНФОРМАЦИИ ПРИ АНАЛИЗЕ ТЕКСТА

1. Понимание «непрямых» способов
выражения смысла (не описанных в
словаре метонимических обозначений,
аллюзий и т.п.)

(«Как будто сам не испытал,
Как режет сталь... » (Евг. Агранович)

В словаре: СХ (сталь) = ВЕЩЕСТВО,
ТВЕРДОЕ

Энц (сталь) = Usage (холодное оружие))



РОЛЬ ЭНЦИКЛОПЕДИЧЕСКОЙ ИНФОРМАЦИИ ПРИ АНАЛИЗЕ ТЕКСТА

2. Подготовка данных для Information extraction

Mesur (аспирин) = грамм; миллиграмм

Mesur (героин) = грамм; килограмм

Mesur (багаж) = килограмм

Mesur (нефть) = баррель

Mesur (вероятность) = %)



3. Разрешение многозначности языковых единиц

пятиметровая доска = > длина;

пятиметровый потолок = > высота;

пятилетний ребенок = > возраст;

пятилетний план = > срок

В словаре:

Param (доска) = длина

Param (потолок) = высота и т.п.

сырой воздух => влажность

сырые овощи => способ приготовления

сырая идея => мера новизны

В словаре:

Param (воздух) = влажность;

Param (овощ) = способ приготовления;

Param (идея) = мера новизны.



5. Описание «хорошо структурируемых» классов

подобно описанию электроизмерительных приборов в работе

Леонтьева Н.Н., Волковыская Е.В.,

Копылова О.В., Молчанова Т.В., Штернова О.А.

Словарь энциклопедических функций и его роль в автоматическом индексировании.

– НТИ. Сер. 2. № 7, 1978

В данной ПО были задействованы функции:

принцип действия, способ

включения/выключения, входной/выходной

объект и др.



6. Верификация текстовых данных

* 31 сентября

В словаре:

Param (сентябрь) = продолжительность;
продолжительность (сентябрь) = (30, «день»)

7. Организация рассуждений

Терапевт N работает по четным утром, по нечетный вечером.

Как N работает 30 мая?

В словаре: Param (ноль) = четность;

Четность (ноль) = +



8.7 Дополнительные лексические связи в словаре и в тексте

(по сравнению с тезаурусными);

дополнительная структуризация связей

(по сравнению с ассоциативной связью)

(давление – манометр;

стакан – стекло

пустыня - саксаул и т.п.)



8.8. Дополнительный формализм для отражения элементов толкования

«Велосипед = двухколесная или трехколесная машина для езды, приводимая в движение ногами»
(МАС)

В словаре:

СХ (велосипед) = УСТРОЙСТВО & СВЯЗАНО &
ДВИЖЕНИЕ

ЭНЦ (велосипед) = {СОСТАВНАЯ_ЧАСТЬ (колесо)

Param (СОСТАВНАЯ_ЧАСТЬ (велосипед) =
количество;

количество (СОСТАВНАЯ_ЧАСТЬ(велосипед)) = (2,
«колесо»); (3, колесо))}.



СПАСИБО ЗА ВНИМАНИЕ!

