

О РАЗРАБОТКЕ МНОГОФУНКЦИОНАЛЬНОГО МНОГОЯЗЫЧНОГО ИНТЕРНЕТ СЕРВИСА НА БАЗЕ МОДЕЛИ ТЮРКСКОЙ МОРФЕМЫ

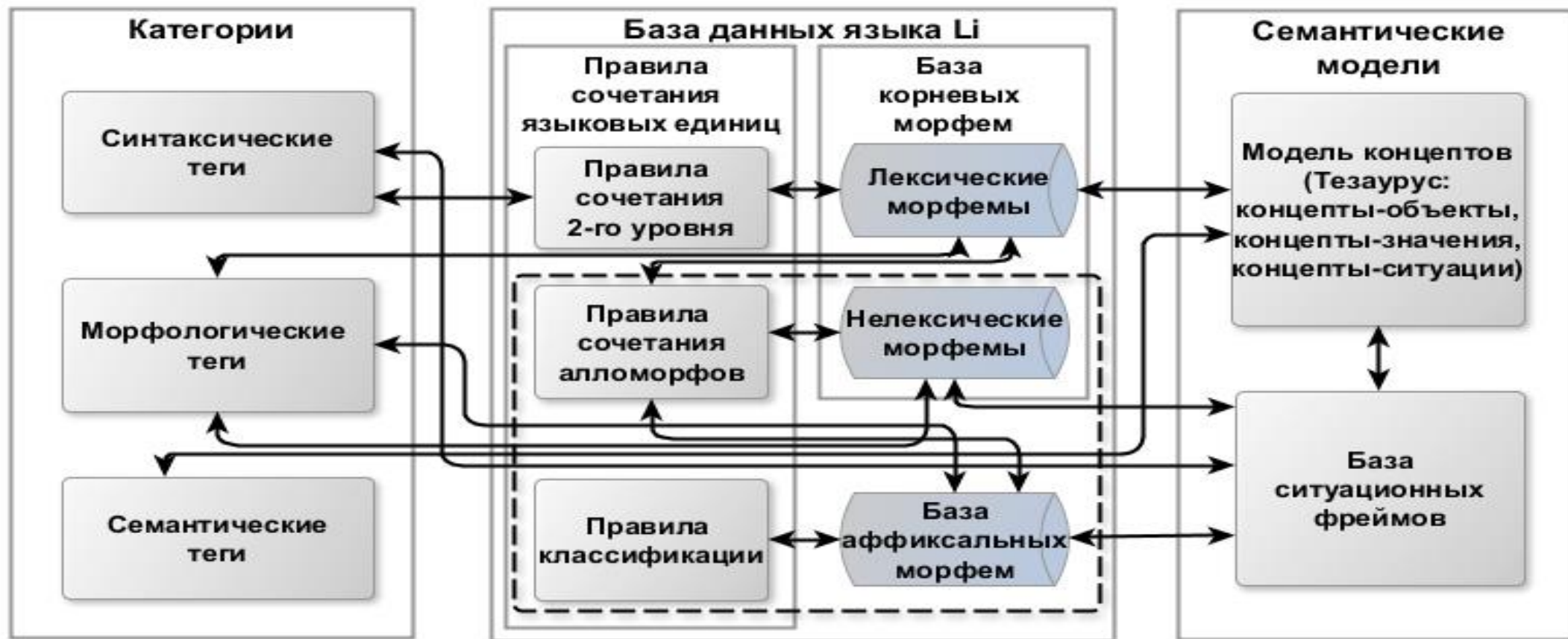
АЛЬМЕНОВА А.Б., ГАТИАТУЛЛИН А.Р., БАШИРОВ А.М.

ТАШКЕНТ-2018

АКТУАЛЬНОСТЬ

- В области разработки систем и технологий по компьютерной обработке тюркских языков **отсутствует реальная интеграция исследований по обработке тюркских языков, происходит дублирование лингвистических моделей и ресурсов, а также программных модулей их обработки, которые на 70–80% являются общими для всех тюркских языков.**
- **Необходимы концептуальные и формальные лингвистические модели, базы данных и знаний, построенные по общим принципам и технологиям для тюркских языков, а также программное обеспечение, созданное на основе этих моделей.**
- **Существующие компьютерные модели тюркских языков не носят интегральный характер.**

Общая архитектура структурно-параметрической функциональной модели



Правила сочетания языковых единиц

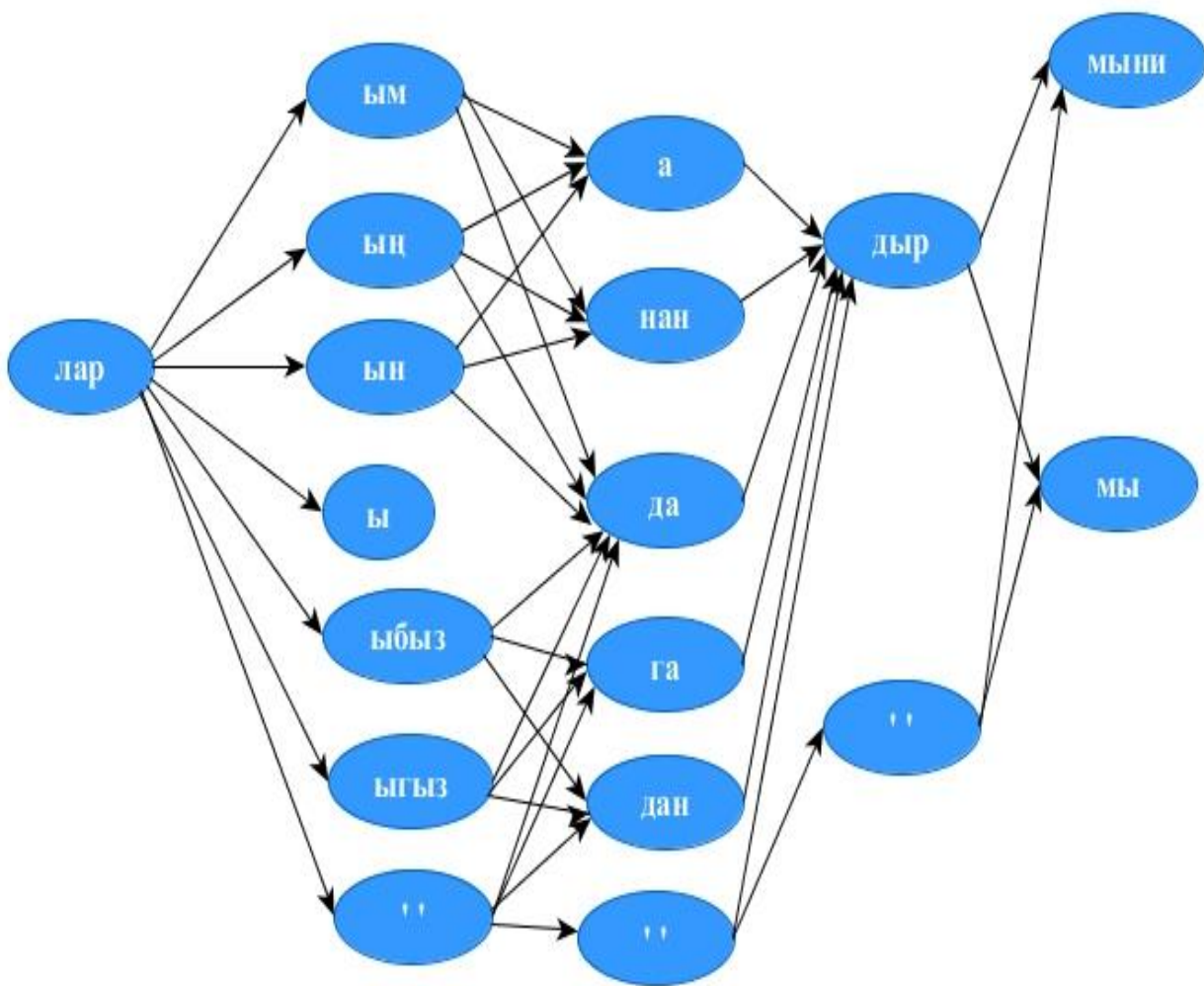
Класс 1 – правила сочетания внутри одной словоформы

Класс 2 – правила сочетания, образующие аналитические формы.

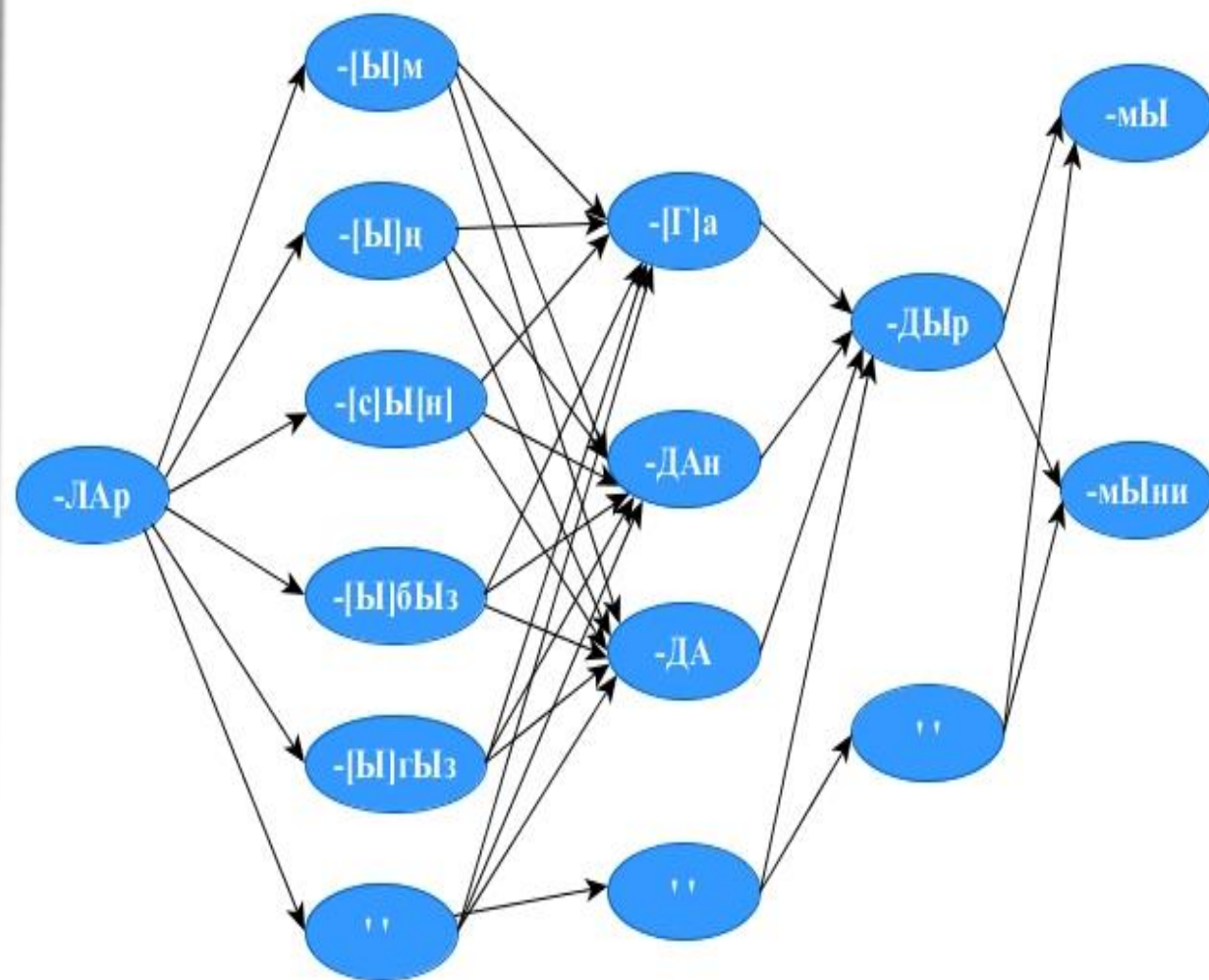
Класс 3 – правила сочетания в предложениях

Граф сочетаний для -Лар морфемы в классе 1

Сочетание алломорфов



Сочетание морфем



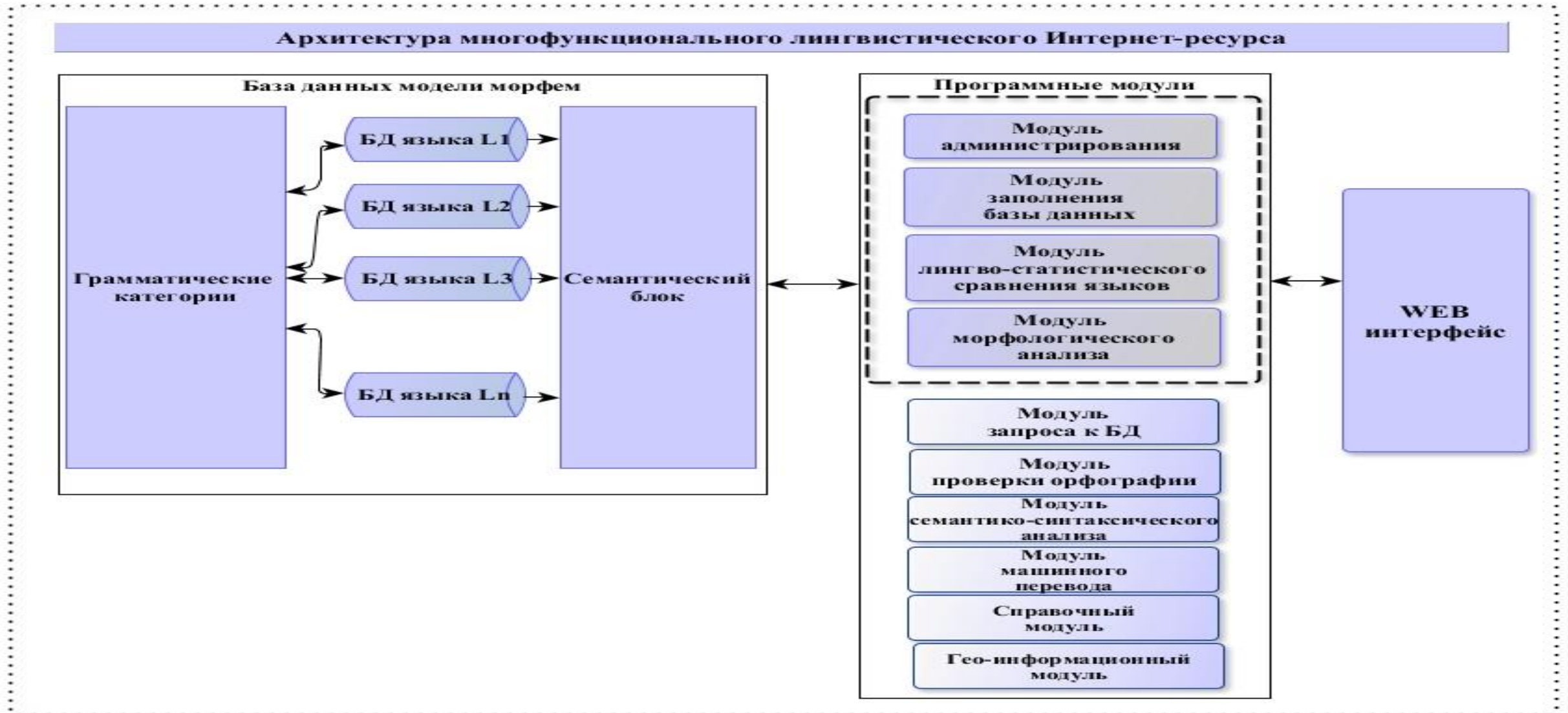
Правила класса 3 – правила сочетания языковых единиц в предложениях

Левая словоформа				Правая словоформа			Пример на ТЯ
Правая морфема	Морфо. Тип	Сем-синт. тип	Приоритет	Морфо. тип	Аффиксы	Роль	
сатин	N	Material	1	N		Object	сатин күлмәк
укытучы	N	Occupation	1	N		Object	укытучы кыз
матур	N	Sign_N	2	N/NP		Object	матур күлмәк
чибәр	N	Sign_N	2	N/NP		Object	чибәр кыз
биш	D	Quant	3	N/NP		Object	биш чибәр кыз
-сыман	SIM_2	Similarity	4	N/NP		Object	фәрештәсыман биш чибәр кыз
-сымак	SIM_3	Similarity	4	N/NP		Object	
-сыз	ATTR_ABES	Without	5	N/NP		Object	жиңсез сатин күлмәк
-лы	ATTR_MUN	With	5	N/NP		Object	жиңле сатин күлмәк
-дагы	ATTR_LOC	Place	6	N/NP		Object	кибеттәге жиңсез биш сатин күлмәк
-ның	GEN	Owner	7	N/NP	POSS.3	Object	абыйның кибеттәге биш сатин күлмәге

Морфологические классы корневых морфем

№	Обозначение	Описание морфологического класса
1	N	класс, который включает в себя корневые морфемы, справа от которых могут следовать все аффиксы именных категорий. Например, аффиксы для класса $N = \{-лАр, -[Ы]м, -[Ы]н, -[с]Ы[н], -[Ы]бЫз, -[Ы]гыз, -[Г]А, -[н]Дан... \}$
2	A	класс, который включает в себя корневые морфемы, справа от которых могут следовать все аффиксы имен прилагательных, которые состоят из всех аффиксов категорий N , а также степени сравнения для прилагательных. Например: аффиксы для класса $A = N + \{-рАк\}$.
3	V	включает в себя корневые морфемы, справа от которых могут следовать все аффиксы, присоединяемые к глагольным основам. Например, аффиксы для класса $V = \{-мА, -ЫЙ, -ДЫ, -ГАН, -АчАк . . . \}$
4	D	корневые морфемы, справа от которых могут следовать аффиксы из категории N и плюс еще аффиксы, образующие формы числительного. Например, аффиксы для класса $D = \{-Ынчы, -[ш]Ар, -ГА, -[н]Дан, -[н]ДА, -ны ... \}$
5	S	корневые морфемы, справа от которых могут следовать только модальные аффиксы. Например, аффиксы для класса $S = \{-мы, -мыни, -Дыр\}$

Архитектура многофункционального лингвистического интернет-сервиса



Структура главного меню интернет-сервиса

адрес: <http://tmm.turklang.tatar/>



Фрагмент интерфейса модуля заполнения базы данных: редактирование идентификаторов

Идентификаторы и алломорфы Правила следования Синтаксический аспект Семантический аспект

-[Ы]М	01.2.002
-[Ы]Ң	01.2.003
-[С]Ы[Н]	01.2.004
-[Ы]быз	01.2.005
-[Ы]гыз	01.2.006
-мын	01.2.008
-сың	01.2.009
-быз	01.2.011
-сыз	01.2.012

Идентификаторы Алломорфы

Обозначение морфемы -[Ы]быз

Цифровой идентификатор 01.2.005

Идентификатор для разметки корпуса POSS.1PL (1st person plural poss)

Название морфологической категории

Типологическое (англ.) 1st person plural possessive

Типологическое (рус.) Притяжательность+ Множественн

Национальное Тартым + куплек сан + 1 зат

Слитность/Раздельность Слитно

Сохранить Добавить и сохранить **Удалить**

СВЯЗЬ КОНЦЕПТОВ И КОРНЕВЫХ МОРФЕМ

Languages

Concepts

ID: 1379
Level: 0
Name: squirrel
NameR: белка

ID: 8434
Level: 0
Name: kopeck
NameR: копейка

Tatar

ID: 01.29137
DictForm: тиен
Root: тиен
MorphonType: 108
Sem1: 1379
Sem2: 8434

Kazakh

ID: 02.1034
DictForm: ақтиін
Root: ақтиін
MorphonType: 110
Sem1: 1379

ID: 02.15630
DictForm: тиін
Root: тиін
MorphonType: 110
Sem1: 1379

ID: 02.15653
DictForm: тиын
Root: тиын
MorphonType: 109
Sem1: 8434

ID: 02.15653
DictForm: копейка
Root: копейка
MorphonType: 101
Sem1: 8434

Turkish

ID: 03.9683
DictForm: sincap
Root: sinca
MorphonType: 113
Sem1: 1379

ID: 03.11563
DictForm: kapik
Root: kapik
MorphonType: 119
Sem1: 8434

Crimean

tatar

ID: 04.4591
DictForm: сансар
Root: сансар
MorphonType: 105
Sem1: 1379

ID: 04.7072
DictForm: капик
Root: капик
MorphonType: 119
Sem1: 8434

Uzbek

ID: 05.2744
DictForm: olmaxon
Root: olmaxon
MorphonType: 102
Sem1: 1379

ID: 05.4497
DictForm: tiyin
Root: tiyin
MorphonType: 102
Sem1: 8434

Концепты WordNet

ID: 29679 Level: 0
Name: entity
NameR: сущность

ID: 29678 Level: 0
Name: physical entity
NameR: физическая сущность

ID: 29680 Level: 0
Name: abstract entity
NameR: абстрактная сущность

ID: 29677 Level: 0
Name: physical object
NameR: физический объект

ID: 2553 Level: 0
Name: matter
NameR: материя

ID: 29676 Level: 0
Name: whole
NameR: целое

ID: 15579 Level: 0
Name: living thing
NameR: живое существо

ID: 29684 Level: 0
Name: natural object
NameR: природный объект

ID: 29685 Level: 0
Name: artefact
NameR: природный объект

ID: 11579 Level: 0
Name: organism
NameR: организм

ID: 11579 Level: 0
Name: biont
NameR: бионт

ID: 13870 Level: 1
Name: plant
NameR: растение

ID: 5795 Level: 1
Name: animal
NameR: животное

ID: 5795 Level: 1
Name: person
NameR: человек

ID: 29212 Level: 2
Name: invertebrate
NameR: беспозвоночные

ID: 27065 Level: 1
Name: chordate
NameR: хордовые

Концепты, отсутствующие в WordNet

18946	Оохотник с беркутом	5609	hunter
26601	Оохотник с кречетом	5609	hunter
27901	Олисогон	5609	hunter
28233	Оохотник на белок	5609	hunter

19044	Обашкир	24866	Russian
20623	Оингуш	24866	Russian
19042	Обалкарец	24866	Russian
19026	Оадыгеец	24866	Russian
19018	Оабазинец	24866	Russian
19055	Обурят	24866	Russian
25903	Охакас	24866	Russian
27721	Отувинец	24866	Russian
21064	Окалмык	24866	Russian

[List of stems](#)[List of concepts](#)[Sequence rules](#)

Concept properties

[Entity](#)[Adjective](#)[Action](#)

Concepts

Name:	cats
Author:	John D.
Visibility Level:	0
Direct hypernym:	feline
Direct hyponym:	domestic cat wild cat

Languages

Tatar:	мәче
Kazakh:	мысық
Kyrgyz:	мышык , пишек
Uzbek:	mushuk
Crimean Tatar:	мышыкъ

Add new concept(s)

Attach existing concept(s)

Concept properties

Concept name

Интерфейс многоязычного морфологического анализатора

ModMorph - Морфологический анализатор для тюркских языков

Файл Анализ Параметры Справка

тиеннарне

<тиеннарне>
1) N:mammal:animal(тиен) + PL (-лар) + ACC (-н[ы])
Synonims:
Crimean Tatar:
сансарларны: N(сансар) + PL (-лар) + ACC (-ны)
Uzbek:
olmaxonlarni: N(olmaxon) + PL (-lar) + ACC (-ni)
Turkish:
sincapları: N(sincap) + PL (-lar) + ACC (-[y]l)
Kazakh:
тиіндерді: N(тиін) + PL (-лар) + ACC (-н[ы]),
ақтиіндерді: N(ақтиін) + PL (-лар) + ACC (-н[ы])
Altay:
тийиндерди: N(тийин) + PL (-лар) + ACC (-ны)

2) N:monetary unit :measure(тиен) + PL (-лар) + ACC (-н[ы])
Synonims:
Uzbek:
tiyinlarni: N(tiyin) + PL (-lar) + ACC (-ni)
Turkish:
kapikkleri: N(kapik) + PL (-lar) + ACC (-[y]l)
Kazakh:
копейкаларды: N(копейка) + PL (-лар) + ACC (-н[ы]),
тиындарды: N(тиын) + PL (-лар) + ACC (-н[ы])

Найдено 1 слов/фраз (100%) за 0.000174407 сек., прибл. скорость 5733.71; не удалось обнаружить сл

Thank you for attention!

Игътибарыгыз өчен рәхмәт!

Назарларыңызга рахмет!

E'tiboringiz uchun rahmat!

Спасибо за внимание!