

Modern Tools For Working With the Tatar Corpus “Tugan Tel”. Module For Morphological Disambiguation

Damir Mukhamedshin
Tatarstan Republic Academy of Sciences
Institute of Applied Semiotics

Morphological Ambiguity

Результаты поиска

Количество результатов: 11, поиск занял 0.002 сек.

Сохранить ▾

11. камалетдинов кэрэзле телефонга ия	и+V+PRES(Й); ия+N+Sg+Nom;	Наил ВАХИТОВ . Мошенник сержант
12. 22 июнь көннеменделеев шаһаренен пионерия урамында әлеге эйберләр	эйбер+N+PL(ЛАр)+Nom; эйберлә+V+FUT_INDF(Ыр); эйберлә+V+PCP_FUT(Ыр);	кант
13. Сержант өчен иң	иң+N+Sg+Nom; иң+V+IMP_SG();	Ың кесәсеннән алынган меңлекләргә алдан ук тамга салынган булган. Наил ВАХИТОВ . Мошенник сержант
14. Чынлап та, иректә Мошенник сержант	ирек+N+Sg+LOC(ДА); ирек+V+CAUS(т)+PRES(Й);	иргәнче, район полициясенең икътисадый жинаятъләргә каршы көрәш бүлегенә мөрәжәгать иткән. Наил ВАХИТОВ .
15. Имеш, телефон, кору жайланмасын ВАХИТОВ . Мошенник сержант	жайла+V+REFL(Ын)+NEG(МА)+JUS_SG(сын); жайлан+V+NEG(МА)+JUS_SG(сын); жайланма+N+Sg+POSS_3SG(Сы)+ACC(ны);	арны сатарга исәп иткән, димәк ки, акчалата бүләкне дә алдан үзләштергән. Наил
16. Уйдырма сержант	уй+V+CAUS(Дыр)+NEG(МА)+IMP_SG(); уйдыр+V+NEG(МА)+IMP_SG(); уйдырма+N+Sg+Nom;	сбатлап та булмый, чөнки аннан алынган элементә чарасы иясенә барып җитмәгән бит. Наил ВАХИТОВ . Мошенник
17. Шул сәбәпле, тикшерү органнарына	ор+V+PCP_PS(Ган)+PL(ЛАр)+POSS_3SG(Сы)+DIR(ГА); орган+N	а туры килә. Наил ВАХИТОВ . Мошенник сержант
18. Моны район прокуратурасы законлы дип таба ,	таба+N+Sg+Nom; таба+POST; тап+V+PRES(Й);	қинаятъ өчен кәшега гаепләү белдереп булмый. Наил ВАХИТОВ . Мошенник сержант
19. getHours () + '& ppr = '+ afReferrer + "'>< \ / scr + 'pr', // - -> казан , http://tat.tatar-inform.ru/	казан+N+Sg+Nom; казан+PROP+Sg+Nom; казан+V+IMP_SG();	йгөл фәхретдинова). "Татар-информ" мәгълүмат агентлыгы. // Источник:
20. Бу якшәмбедә илебезнең авыл хужалыгы һәм эшкәртү сәнәгате хезмәтчәннәре һөнәри бәйрәмнәрен билгеләп http://tat.tatar-inform.ru/	билге+N+DISTR(лап); билгелә+V+ADVV_ACC(Ып);	тлыгы. // Источник:

Users' Roles

- ▶ Guest
- ▶ Annotator
- ▶ Main Annotator
- ▶ Administrator

Document Choosing

Список документов, доступных для снятия многозначности

("Татар-информ" мәгълүмат агентлыгы) [1] [Фреймы]

Мошенник сержант (Наил ВАХИТОВ) [2] [Фреймы]

("Татар-информ" мәгълүмат агентлыгы) [3] [Фреймы]

Сембер татар малтабарлары милләттәш депутатлар белән очраштылар (айрат ибраһим) [5] [Фреймы]

("Татар-информ" мәгълүмат агентлыгы) [7] [Фреймы]

Казанда бал ярминкәсе ачылды (ленар мөхәммәдиев) [8] [Фреймы]

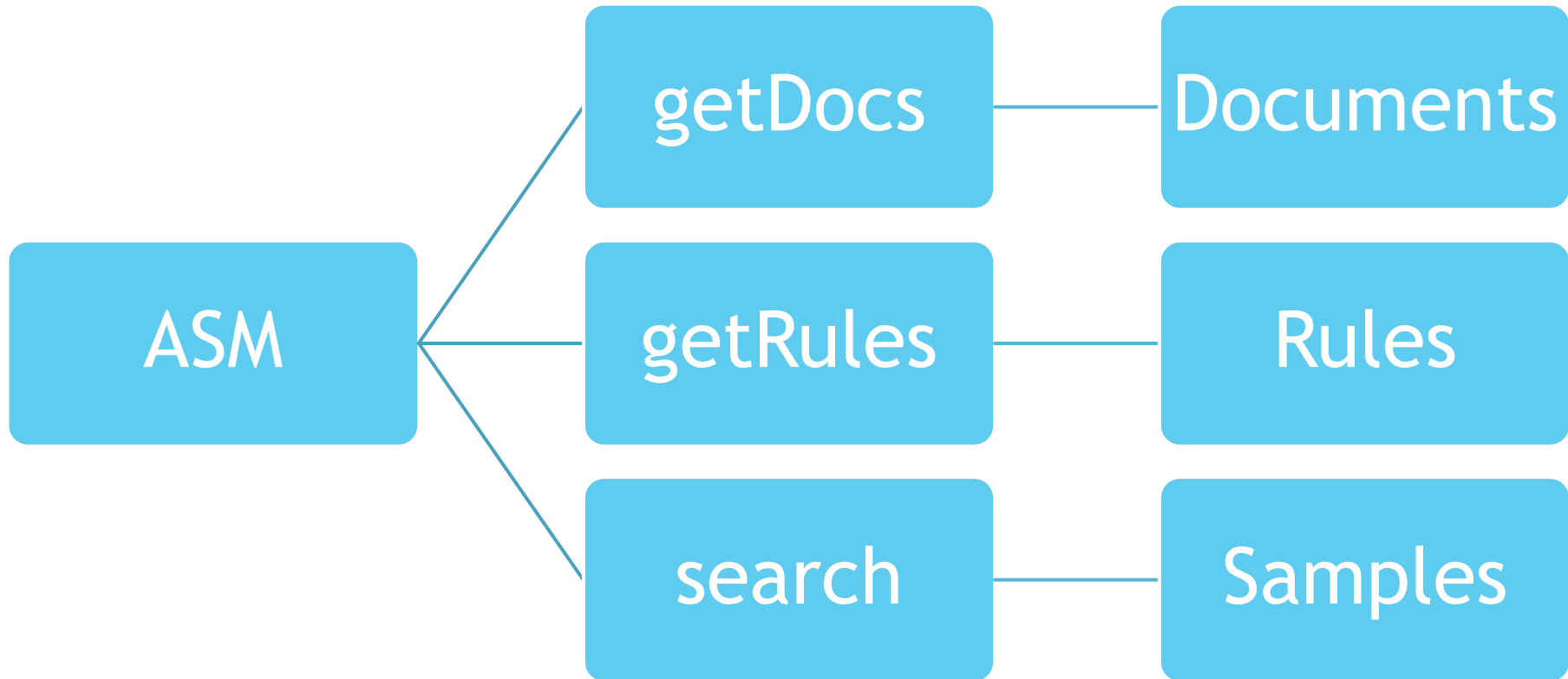
("Татар-информ" мәгълүмат агентлыгы) [9] [Фреймы]

Сыйныфтан тыш чаралар да тәрбияли (Флүзә РӘШИТОВА) [10] [Фреймы]

Бала-чага сугышса да, Президентны борчыйбыз... (Илдар Фазлетдинов) [11] [Фреймы]

Казанда да картлар йорты ачылачак (наил алан) [12] [Фреймы]

AmbiguitySamplesModel



Disambiguation Interface

Примеры

Осталось примеров: 8057

Чараны **Сәламәтлек** саклау министрлыгы , тр традицион медицина буенча белгечләре

чара+N+Sg+ACC(ны) сәламәт+Adj+NMLZ(лык)+Sg+Nom сәламәт+Adj+PSBL(лык) белгеч+N+PL(лар)+POSS_3SG(сы)+Nom

Введите другой разбор Введите другой разбор Введите другой разбор

Пропустить Удалить Ошибка! Пропустить Удалить Ошибка! Пропустить Удалить Ошибка!

оешмасы Кытай фәннәр академиясенең Традицион медицинаның

оешма+N+Sg+POSS_3SG(сы)+Nom кытай+N+Sg+Nom академия+N+Sg+POSS_3SG(сы)+GEN(ның) медицина+N+Sg+GEN(ның)

Введите другой разбор Введите другой разбор Введите другой разбор Введите другой разбор

Пропустить Удалить Ошибка! Пропустить Удалить Ошибка! Пропустить Удалить Ошибка! Пропустить Удалить Ошибка!

нигез теориясе институты , Казандагы Традицион кытай медицинасы үзәге булышыгында

нигез+N+Sg+Nom казан+N+Sg+ATTR_LOC(дагы) казан+PROP+ATTR_LOC(дагы)

Введите другой разбор Введите другой разбор

Пропустить Удалить Ошибка! Пропустить Удалить Ошибка!

Казан дәүләт медицина академиясе оештырды

казан+N+Sg+Nom ое+V+RAR_2(ыштыр)+PST_DEF(ды) казан+PROP+Sg+Nom ое+V+RECP(ыш)+CAUS(дыр)+PST_DEF(ды) казан+V+IMP(Sg) оеш+V+CAUS(дыр)+PST_DEF(ды) оештыр+V+PST_DEF(ды)

Введите другой разбор Введите другой разбор

Пропустить Удалить Ошибка! Пропустить Удалить Ошибка!

Традицион медицина жәмгыять тарафыннан тупланган һәм буыннан буынга

Бүген ул тагын да популяррак була бара .

Сохранить контекст

AmbiguityTrackingModel

- ▶ Working - used to analyze the useful time spent by the user to resolve morphological ambiguity
- ▶ Choose - the action of choosing the correct option from the list in the user's opinion
- ▶ Miss - the action of sample missing
- ▶ Remove - the action of sample removing
- ▶ Correct - the action of sample correction
- ▶ Copy, Cut, Paste - actions with clipboard
- ▶ Blur, Focus - actions with browser window

Disambiguation Interface

Примеры

Осталось примеров: 8057

Чараны **Сәламәтлек** саклау министрлыгы , тр традицион медицина буенча белгечләре

чара+N+Sg+ACC(ны) | сәламәт+Adj+NMLZ(лык)+Sg+Nom | белгеч+N+PL(лар)+POSS_3SG(сы)+Nom

Введите другой разбор | Введите другой разбор | Введите другой разбор

Пропустить Удалить Ошибка!

оешмасы Кытай фәннәр академиясенең Традицион медицинаның

оешма+N+Sg+POSS_3SG(сы)+Nom | кытай+N+Sg+Nom | академия+N+Sg+POSS_3SG(сы)+GEN(ның) | медицина+N+Sg+GEN(ның)

Введите другой разбор | Введите другой разбор | Введите другой разбор | Введите другой разбор

Пропустить Удалить Ошибка!

нигез теориясе институты , Казандагы Традицион кытай медицинасы үзәге булышыгында

нигез+N+Sg+Nom | казан+N+Sg+ATTR_LOC(дагы) | казан+PROP+ATTR_LOC(дагы) | оештырды

Введите другой разбор | Введите другой разбор | Введите другой разбор

Пропустить Удалить Ошибка!

Казан дәүләт медицина академиясе оештырды

казан+N+Sg+Nom | казан+PROP+Sg+Nom | казан+V+IMP(Sg) | ое+V+RAR_2(ыштыр)+PST_DEF(ды) | ое+V+RECP(ыш)+CAUS(дыр)+PST_DEF(ды) | оеш+V+CAUS(дыр)+PST_DEF(ды) | оештыр+V+PST_DEF(ды)

Введите другой разбор | Введите другой разбор

Пропустить Удалить Ошибка!

Традицион медицина жәмгыять тарафыннан тупланган ләм буыннан буынга

Бүген ул тагын да популяррак була бара .

Сохранить контекст

Rules for Samples Generating

ВГ-октябрь2019

Название фрейма
ВГ-октябрь2019

Поиск по
Словоформе

Формула
*

Ограничение (0 - без ограничений)
0

Включить словоформы без многозначности (сплошное редактирование)

Подкорпус
Подкорпус разрешения многозначности

Документ
2_1551_1_1.txt
2_1552_1_1.txt
2_1553_1_1.txt
2_1554_1_1.txt
2_1555_1_1.txt
2_1556_1_1.txt
2_1557_1_1.txt
2_1558_1_1.txt
2_1559_1_1.txt

Группа пользователей
Все группы

Внимание! Обновление фрейма приведет к потере всех изменений! Обновляйте фрейм только если допустили ошибку, или никто еще не приступал к разрешению этого фрейма.

Внимание! Добавление примеров может занять продолжительное время. Особенно для списков документов.
Даже если сервер вернул ошибку 504, не обновляйте страницу, примеры будут добавлены в фоне!

Изменить фрейм и обновить примеры

Users' Statistics

Туган Тел
НАЦИОНАЛЬНЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

Рус
Тат
Eng

Разрешение многозначности

Аннотаторы

Проверка

Подбор примеров

Статистика

Статистика пользователей

Период:

дд.мм.гггг

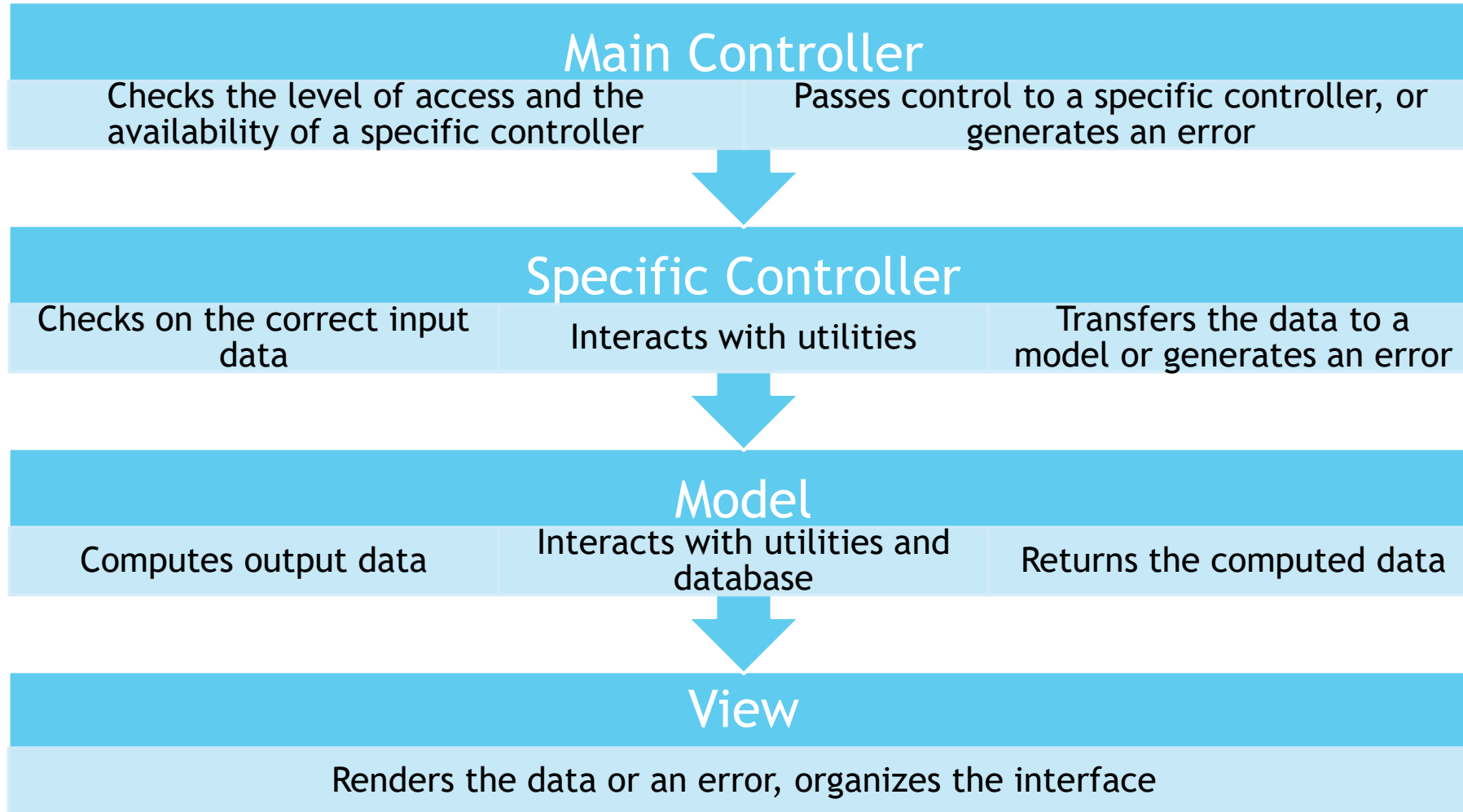
-

дд.мм.гггг

Найти

ID	Имя	Время	Примеры			Символы	
			Выбранные	Проверенные	Пропущенные	Выбранные	Проверенные
1	Дамир	00:52:47	33	6	10	8736	2546
2	Тест	00:00:00	0	0	0	0	0
3	Тест	00:00:00	20	0	4	3270	0
4	Пользователь	00:00:00	0	0	0	0	0
5	Булат Хакимов	00:12:18	0	0	0	0	0
6	Вилюза Гафарова	00:14:39	41	1	2	4526	183
7	Лилия Гатина	00:00:00	0	0	0	0	0
8	Санифа Габбасова	00:00:00	0	0	0	0	0
9	Айсылу Мухаметзянова	00:00:00	0	0	0	0	0
10	Резеда Забирова	00:00:00	0	0	0	0	0

Module Architecture



Module Database

Column name	Data type	Description
id	INT (AUTO-INCREMENT)	Transaction unique identifier
user_id	INT	User identifier
sample_id	INT	Sample identifier
chosen	SMALLINT	Chosen sample number
datetime	DATETIME	Date and time of action
status	ENUM ('chosen', 'checked', 'missed', 'removed', 'checked_removed')	Action taken on the sample
corrected	VARCHAR	User's manual corrections

Conclusion

- ▶ User-friendly interface
- ▶ «Infinity» samples
- ▶ Sample missing possibility
- ▶ Selection by documents, samples, missed samples
- ▶ Automatic tracking of time and volume of work
- ▶ Correcting wordforms and morphological attributes
- ▶ Continuous editing
- ▶ Tracking data can be used for automatic morphological disambiguation
- ▶ Can be used for corpora of other languages

Thank you for attention!

damirmuh@gmail.com
<https://t.me/damirmuh>