

The Main Results of Implementation of
The State Program "Conservation, Research and
Development of the State Languages of the Republic of
Tatarstan and other Languages in the Republic of Tatarstan
for 2014 - 2020 years"

Rinat Gilmullin

Institute of Applied Semiotics of TAS, Kazan, Tatarstan

rinatgilmullin@gmail.com

The Main Directions of the State Program

- Tatar Localization of the Information Technologies
- Mobile applications and Web-Services
- Speech-to-Text and Text-to-Speech systems for the Tatar
- Tatar Nacional Corpus
- Russian-Tatar MT

Tatar Localization of the Information Technologies

- Support of Tatar Language in Windows (from Windows NT to Windows 10)

Language Interface Pack for Windows 7, Vista, 8, 8.1	Language Interface Pack for Microsoft Office 2007, 2010, 2013, 2016
https://www.microsoft.com/tt-ru/download/confirmation.aspx?id=17036 https://www.microsoft.com/tt-ru/download/confirmation.aspx?id=4903	https://www.microsoft.com/tt-ru/download/confirmation.aspx?id=20385 https://www.microsoft.com/tt-ru/download/confirmation.aspx?id=6804 https://www.microsoft.com/tt-RU/download/details.aspx?id=36528

Tatar Localization of the Information Technologies

- Tatar localization of the **Sailfish Mobile OS**

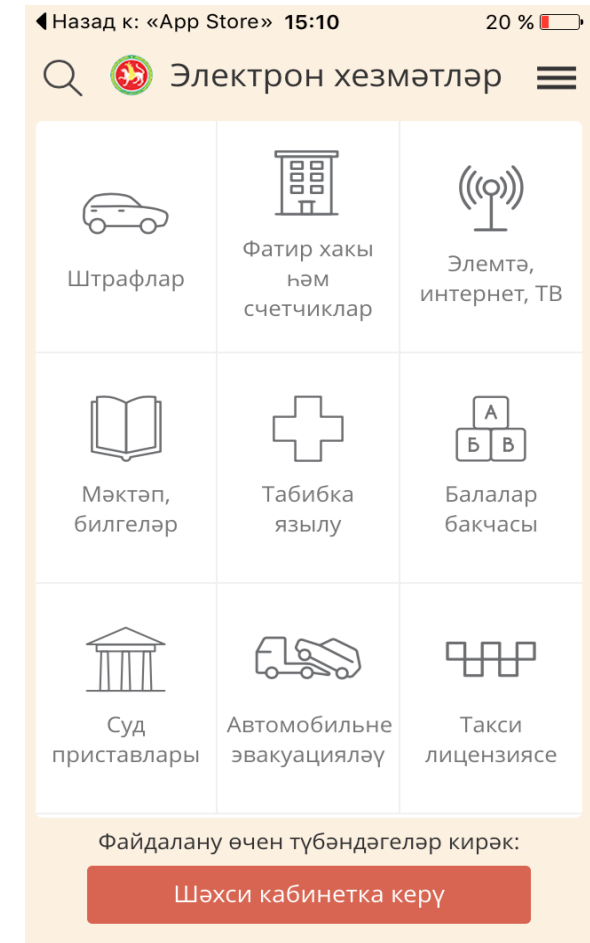
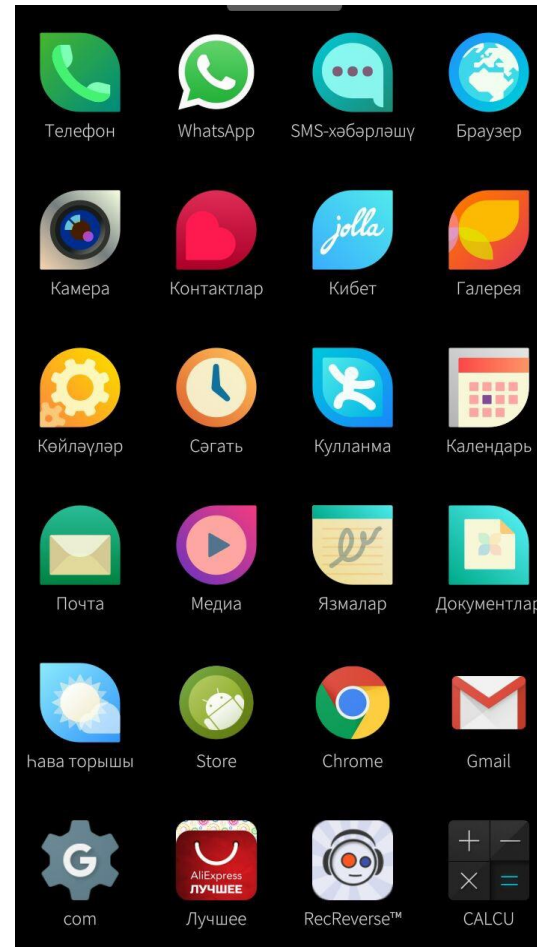
<http://mic.tatarstan.ru/rus/index.htm/news/948912.htm>

Jointly with the company "Open mobile platform"

- Tatar localization of the «Mobile app «Uslugi RT»»

<https://uslugi.tatarstan.ru/start/mobile-apps>

✓ more 20 000 terms and phrase



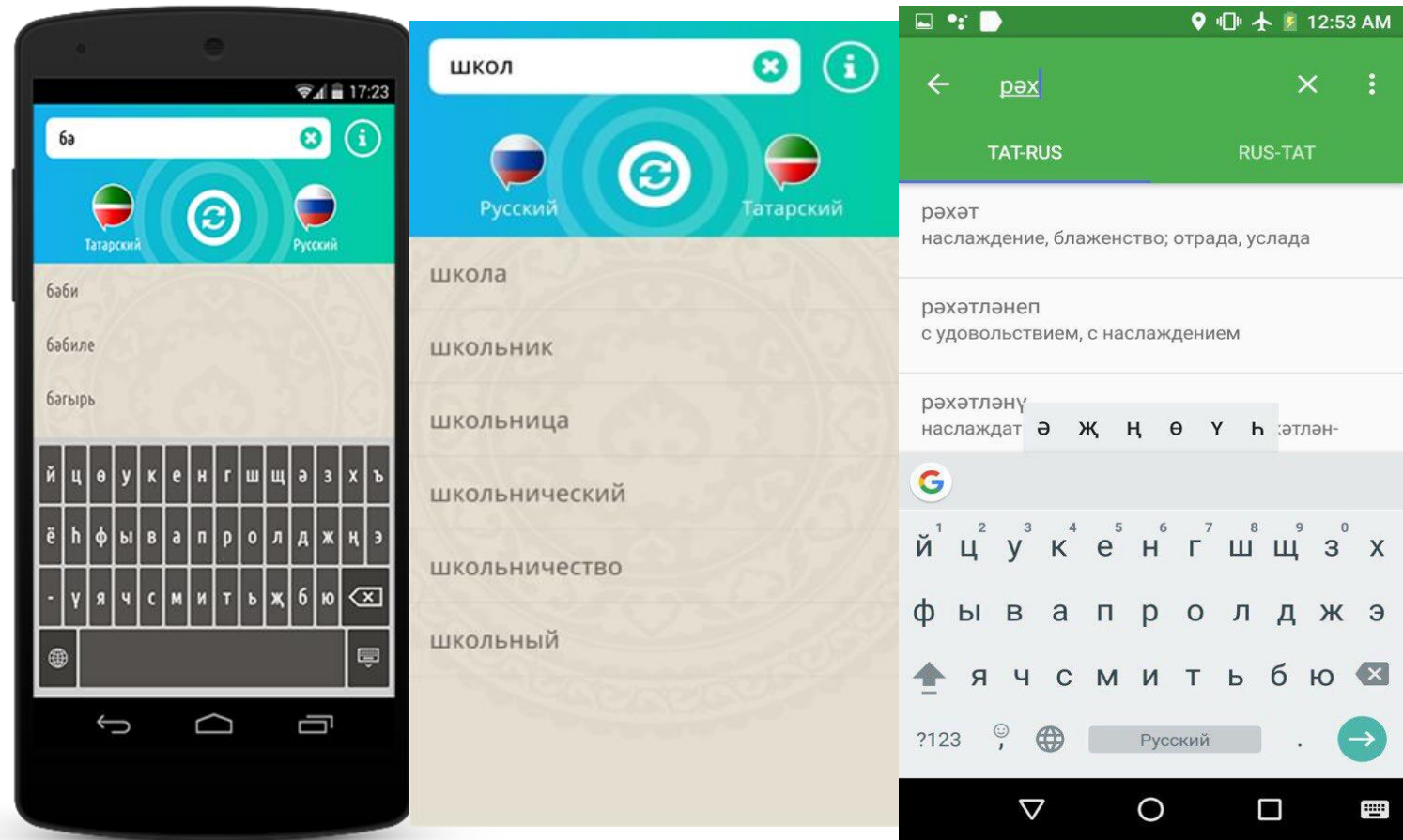
Mobile apps and Web-Services

- TatDict (Ru)

more 60,000 words

App Store, Google Play

<https://play.google.com/store/apps/details?id=com.aimstd.TatDict.Ru&hl=ru>



MTS

22:40

8 %

Словарь ИТ терминов



Институт прикладной семиотики
Академия наук Республики Татарстан



MTS 22:44 6 %

ИТ-термины Значение термина

access point

керү ноктасы, тоташу ноктасы

точка доступа

(... патне) кәмелли төл

чыбыксыз челтәр клиенты өчен
концентратор ролен уйнаучы
һәм кабельле локаль челтәргә
тоташуны тәэмин итүче
програм-аппарат җайланма

Dictionary of IT-Terms

- ✓ description of English terms in Russian, Tatar and Chuvash languages
- ✓ more 6,000 terms

Date of publication in App Store: 24.03.2017

Downloads: 2783

1 апр. 2015 — 18 окт. 2017



8,265 Показы



367 Просмотры страницы продукта



2,783 Копии приложения



0 Встроенные покупки



\$0 Продажи



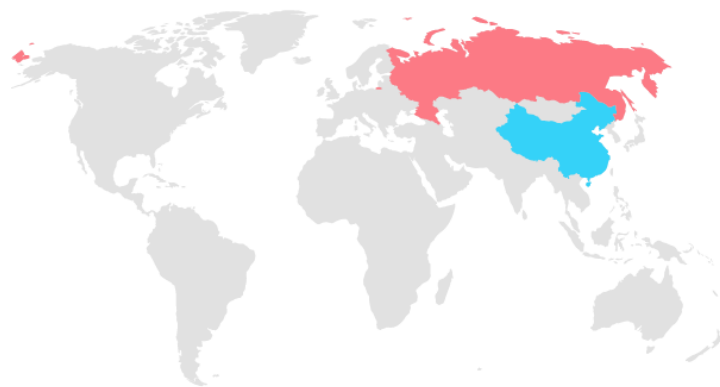
41 Сеансы Получено согласие



1 (в среднем за месяц) Активные устройства Получено согласие



0 Сбои Получено согласие



Копии приложения по странам



http://itdictionary.tatsoft.tatar

Dictionary of IT-Terms

Q computer

Английский-татарский

Искать

Английский-татарский

Английский-русский

Английский-чувашский

Татарский-английский

Татарский-русский

Татарский-чувашский

Русский-английский

Русский-татарский

Русский-чувашский

Чувашский-английский

Чувашский-татарский

Чувашский-русский

классы
кулланыла

Результаты поиска терминов по запросу

computer

rus компьютер, электронная вычислительная машина, ЭВМ

tat компьютер, санак, кампир, электрон хисаплау машинасы, электрон ис

chv компьютер, электронла шутлав машини, ЭШМ

мәгълүмат эшкәртүне автоматлаштыру өчен билгеләнгән жайланма яки була: цифр исәпләү машиналары һәм аналог исәпләү машиналары. Хәзергә

ANACOM (ANALog COMputer)

tat кара: analog computer

analog computer

rus аналоговая вычислительная машина (АВМ)

tat аналог хисап машинасы (АХМ), аналог компьютер

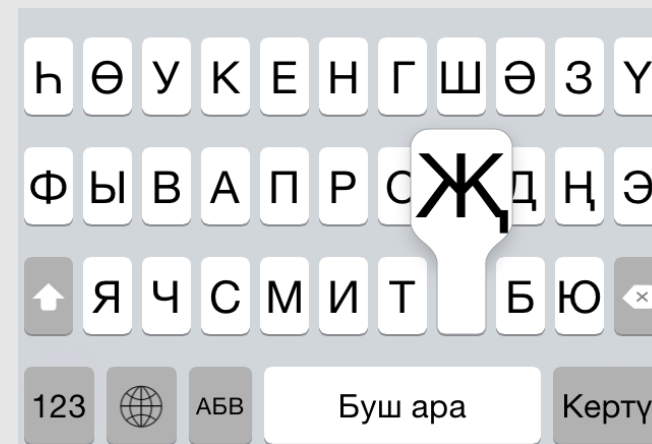
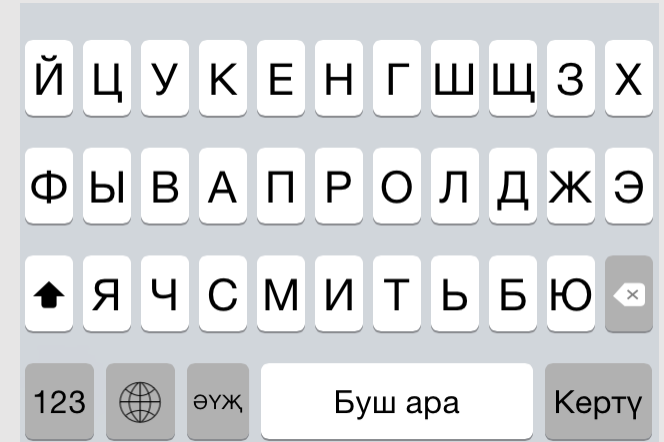
chv аналогла шутлав машини (АШМ)

өзлексез үзгәрәп торучы физик зурлыклар рәвешендә күрсәтелгән мәгълүматне эшкәртүче хисап машинасы. Андый зурлыклар булып, мәсәлән, почмакча һәм сызыкча күчешләр, электр көчәнеше, ток зурлыгы һ.б. торырга мөмкин. Ул зурлыклар математик үзгәрешлеләргә үзлештерелә (үзгәртелә) ала. Аналог зурлыкларны үлчәп, аларны механик яки башка юл белән кушарга, тапкырларга, интегралларга һ.б. гамәлләр башкарырга була. Электрон АХМнар киң таралган. Аларда мәгълүмат электр көчәнеше рәвешендә күрсәтелә һәм эшкәртелә. Математик мәсьәләне чишү өчен электр чылбыры элементларыннан махсус схема жыела. Аның чыгышында көчәнеш башлангыч математик мәсьәлә чишелеше канәгатьләндерергә тиешле тигезләмәләрне канәгатьләндерә. Шулай итеп, көчәнеш эзләнгән математик үзгәрешленең аналогы булып тора. Схема параметрларын билгеле бер вакыт аралыгында үзгәртәп, шул вакыт аралыгында үзгәрүче чишелешләр табып була. Шуңа күрә физик вакыт андый АХМда аналог үзгәрешле булып тора. АХМның төп житешсезлекләре – төгәллеге һәм универсальлеге аз булу. АХМнар башлыча төрле идарә системаларында һәм мәгълүмат жыю һәм эшкәртү системаларында үлчәү жайланмалары буларак кулланылалар.

Синонимы - analog machine (чагыштыр: digital computer)

Tatar keyboard for iOS

- ✓ 12,000 downloading
- ✓ 3 different keyboard layout



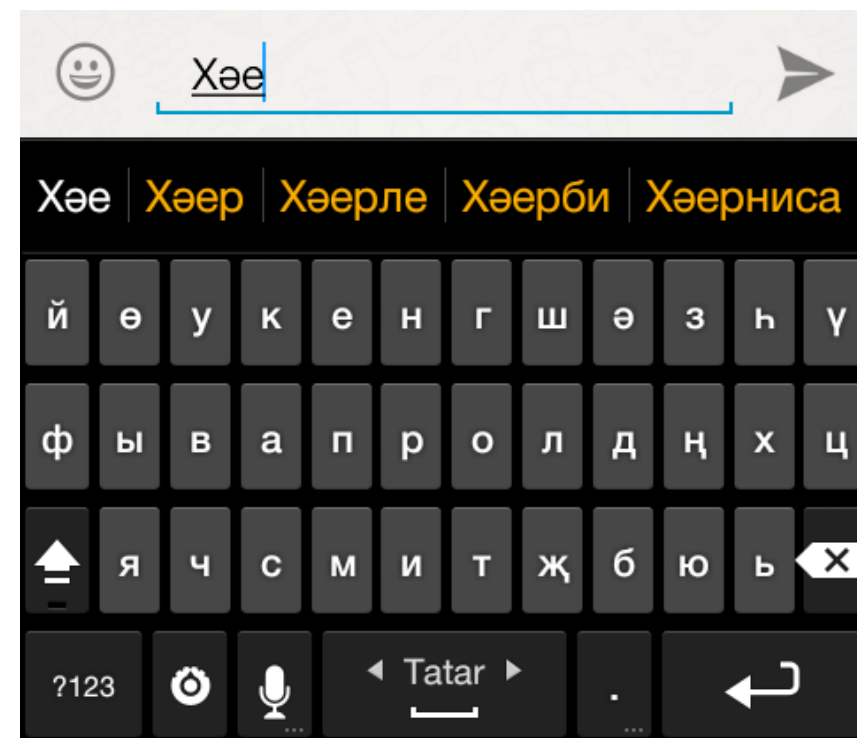
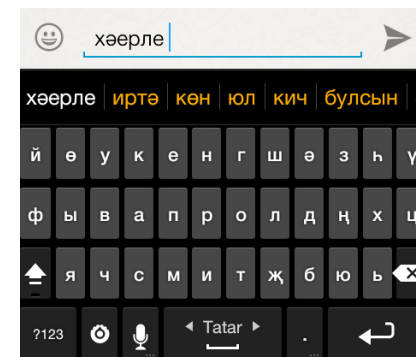
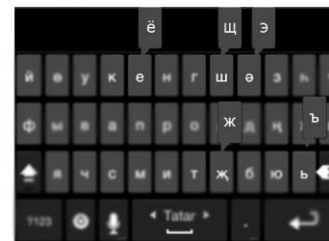
Tatar keyboard for iOS

Диаграмма	Страна	Копии приложения
<input checked="" type="checkbox"/>	Россия	11,910
<input type="checkbox"/>	Казахстан	89
<input type="checkbox"/>	США	76
<input type="checkbox"/>	Украина	18
<input type="checkbox"/>	Турция	16
<input type="checkbox"/>	Кыргызстан	15
<input type="checkbox"/>	Германия	14
<input type="checkbox"/>	Китай	13
<input type="checkbox"/>	Канада	12
<input type="checkbox"/>	Узбекистан	12
<input type="checkbox"/>	ОАЭ	11
<input type="checkbox"/>	Беларусь	7
<input type="checkbox"/>	Таджикистан	7
<input type="checkbox"/>	Япония	6
<input type="checkbox"/>	Великобритания	5
<input type="checkbox"/>	Италия	5
<input type="checkbox"/>	Чехия	5

Тиз.яз for Android

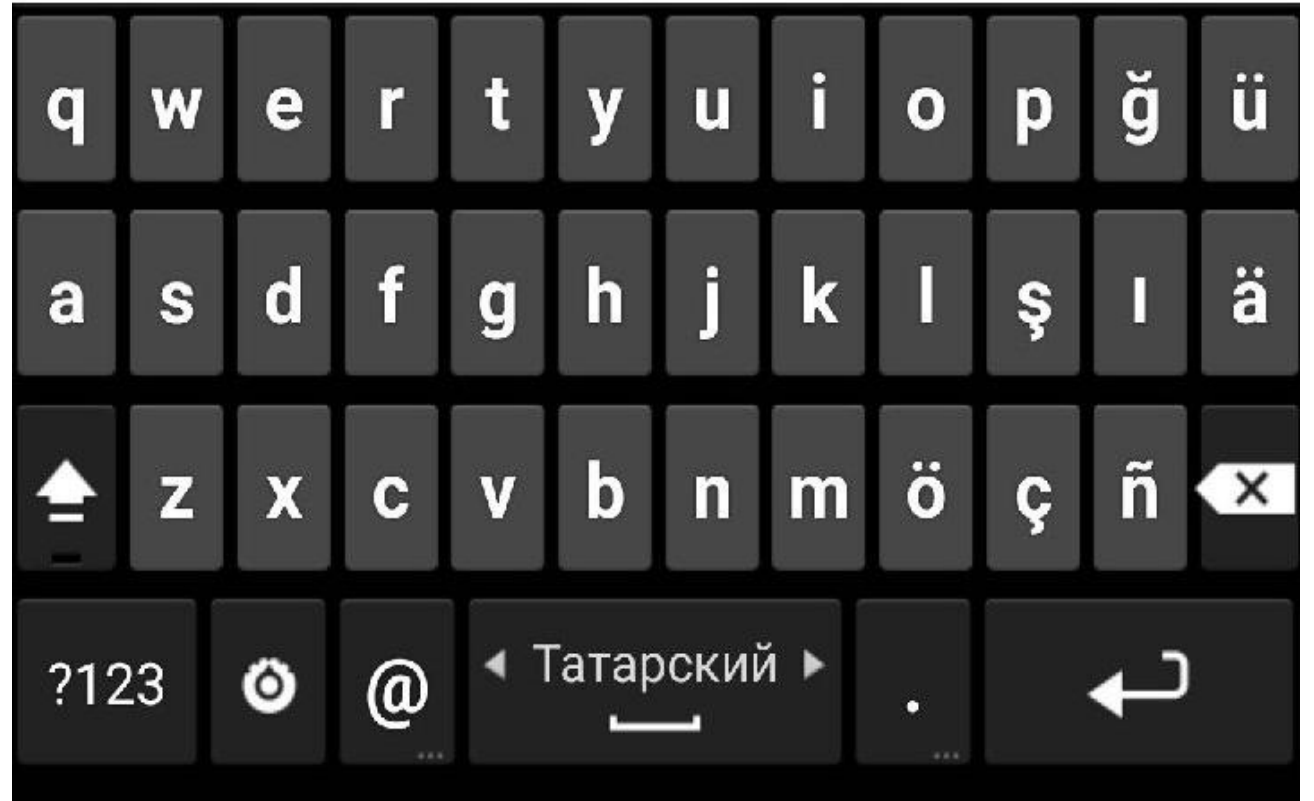
- ✓ 48,000 downloading (since December of 2014)
- ✓ Russia (98%), Kazakhstan (1%), Turkmenistan (0,5%), Uzbekistan (0,5%)
- ✓ Predictive input

Удерживайте клавиши, чтобы
выбрать
дополнительные буквы:



Tiz.yaz on Latin Alphabet

- ✓ Data of Publication: 4.10.2017
- ✓ 11 downloading:
Russia (90%), USA(5%), Germany(5%)



Mobile apps and Web-Services

➤ <http://tatsoft.tatar>

Russian-Tatar MT

- The support system of professional translation SmartCAT
<https://www.smartcat.ai/>
- Open Internet service with the Russian-Tatar machine translation system:
- <https://translate.yandex.ru>

Development of the Linguistic resources for MT

- Cooperation with **Yandex Company** on the development of the system of Tatar-Russian machine translation: <https://translate.yandex.ru/>
- The following resources are prepared and transferred to Yandex:
 - ✓ **Russian-Tatar parallel corpus - the volume of 17,6 million word forms** (несколько тысяч страниц оцифровки и работы по выравниванию)
 - ✓ **Tatar texts - the volume of 40 million word forms**
 - ✓ **Module of the Tatar morphological analysis**
- Cooperation with the **ABBYY LS Company** on
 - ✓ **maintenance of the system for professional translation SmartCAT**
 - ✓ **development of the Linguistic resources for MT**

Development of the Linguistic resources for MT (2017)

- Русско-татарский параллельный корпус ~ 30 млн. словоформ (при помощи краудсорсинговой платформы на базе площадки «Сэлэт»)
- Двухязычные (русско-татарские) специализированные словари фамилий, имен, отчеств, государств, субъектов РФ, районов РТ, населенных пунктов, гражданств, национальностей в общем объеме НЕ МЕНЕЕ 70 000 словарных пар.
- Словарь грамматических основ татарского языка в объеме НЕ МЕНЕЕ 25 000 основ.
- Грамматические таблицы.
 - БД правил для анализа и выделения именных групп в татарском языке на основе электронного корпуса татарского языка. Объем электронного корпуса составляет НЕ МЕНЕЕ 80 млн. словоформ.
 - Сопоставительная таблица именных групп татарского и русского языков на основе русско-татарского параллельного корпуса. Объем используемого русско-татарского параллельного корпуса составляет НЕ МЕНЕЕ 10 млн. словоформ.

Tatar National Corpus “Tugan Tel” (TNC)

✓ <http://tugantel.tatar/>

The total volume of the annotated corpus - 128 million word forms.

TNC includes also a search platform, a software package for the linguistic statistical study of the Language Corpus and the DataBase.

THANK YOU