



Марийский язык в киберпространстве: достижения и проблемы

Андрей Чемышев,
Марийский научно-исследовательский
институт языка, литературы и истории им.
В.М. Васильева

- 
1. Раскладки клавиатуры и символы
 2. Электронные словари
 3. Системы проверки правописания
 4. Машинные переводчики
 5. Синтезаторы речи
 6. Распознаватели речи
 7. ...

1. Раскладки клавиатуры и символы:



Символы:

— регулярный
алфавит:

Ää, Нн, Öö, Ўў, Ыы

Unicode



Символы:

— исторические:

ЎЎ = Ў + ~

не Unicode



Символы:

— Транскрипции:

...



Раскладки клавиатуры:

- Windows
- Linux
- Mac OS
- Android
- iOS

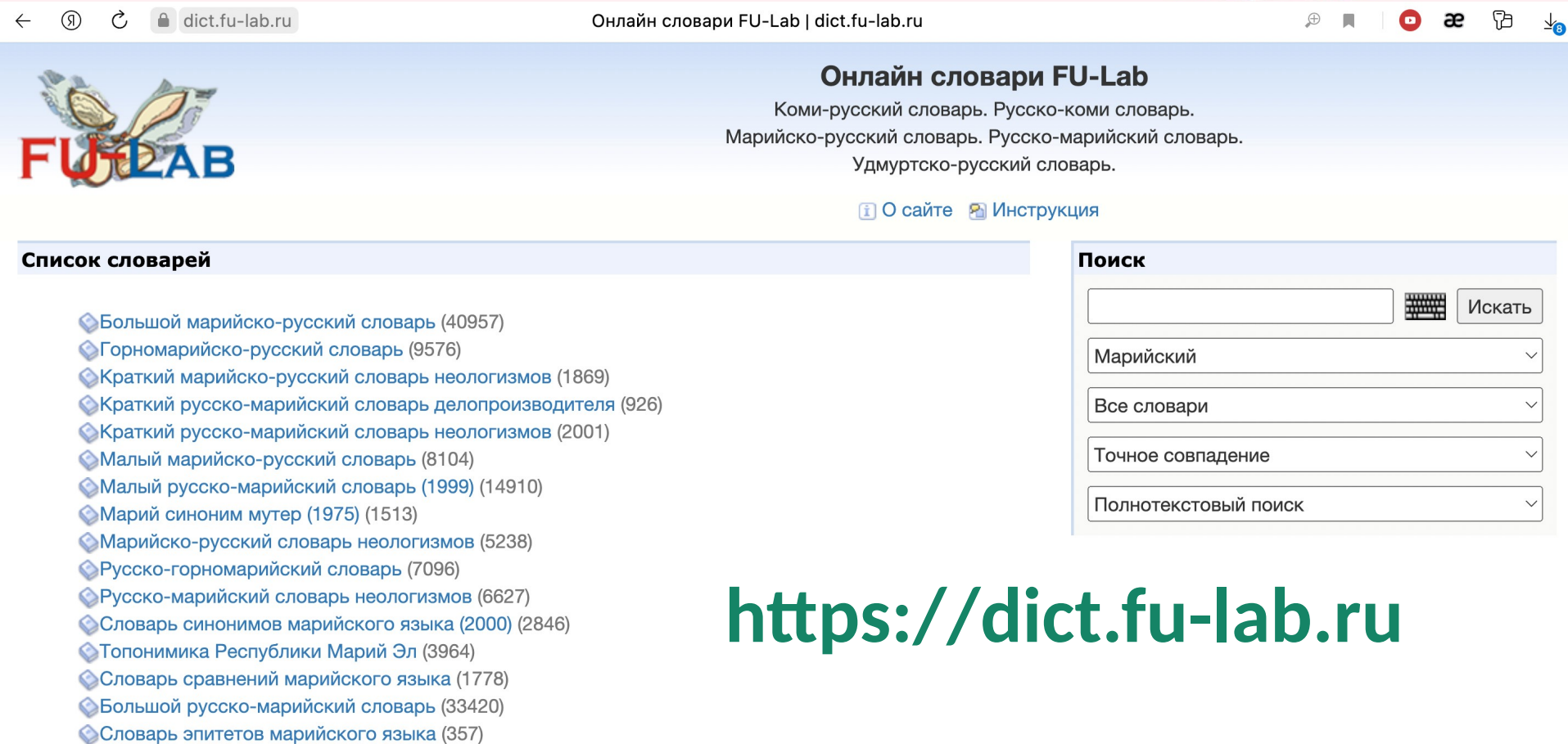


Раскладки клавиатуры:

- Windows
- Linux
- Mac OS
- Android
- iOS
- virtual keyboard ?



2. Электронные словари:



The screenshot shows the homepage of the FU-Lab online dictionaries. The header features a decorative red banner with white floral patterns. Below the banner, the website's title "Онлайн словари FU-Lab" is displayed, followed by a list of available dictionaries: Коми-русский, Русско-коми, Марийско-русский, Русско-марийский, and Удмуртско-русский. A search bar is located on the right side of the page, and a list of dictionaries is on the left. The URL "https://dict.fu-lab.ru" is prominently displayed at the bottom.

← Я ↻ dict.fu-lab.ru Онлайн словари FU-Lab | dict.fu-lab.ru 🔍 📄 📺 ⌂ ⚙️ ⬇️

Онлайн словари FU-Lab

Коми-русский словарь. Русско-коми словарь.
Марийско-русский словарь. Русско-марийский словарь.
Удмуртско-русский словарь.

[О сайте](#) [Инструкция](#)

Список словарей

- ❖ Большой марийско-русский словарь (40957)
- ❖ Горномарийско-русский словарь (9576)
- ❖ Краткий марийско-русский словарь неологизмов (1869)
- ❖ Краткий русско-марийский словарь делопроизводителя (926)
- ❖ Краткий русско-марийский словарь неологизмов (2001)
- ❖ Малый марийско-русский словарь (8104)
- ❖ Малый русско-марийский словарь (1999) (14910)
- ❖ Марий синоним мутер (1975) (1513)
- ❖ Марийско-русский словарь неологизмов (5238)
- ❖ Русско-горномарийский словарь (7096)
- ❖ Русско-марийский словарь неологизмов (6627)
- ❖ Словарь синонимов марийского языка (2000) (2846)
- ❖ Топонимика Республики Марий Эл (3964)
- ❖ Словарь сравнений марийского языка (1778)
- ❖ Большой русско-марийский словарь (33420)
- ❖ Словарь эпитетов марийского языка (357)

Поиск



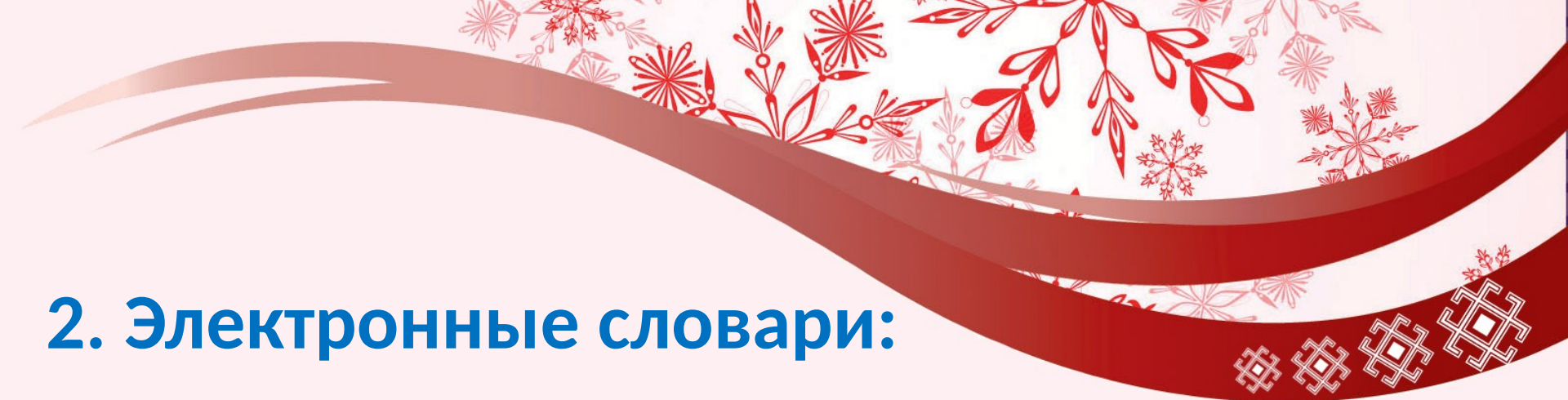
Марийский ▾

Все словари ▾

Точное совпадение ▾

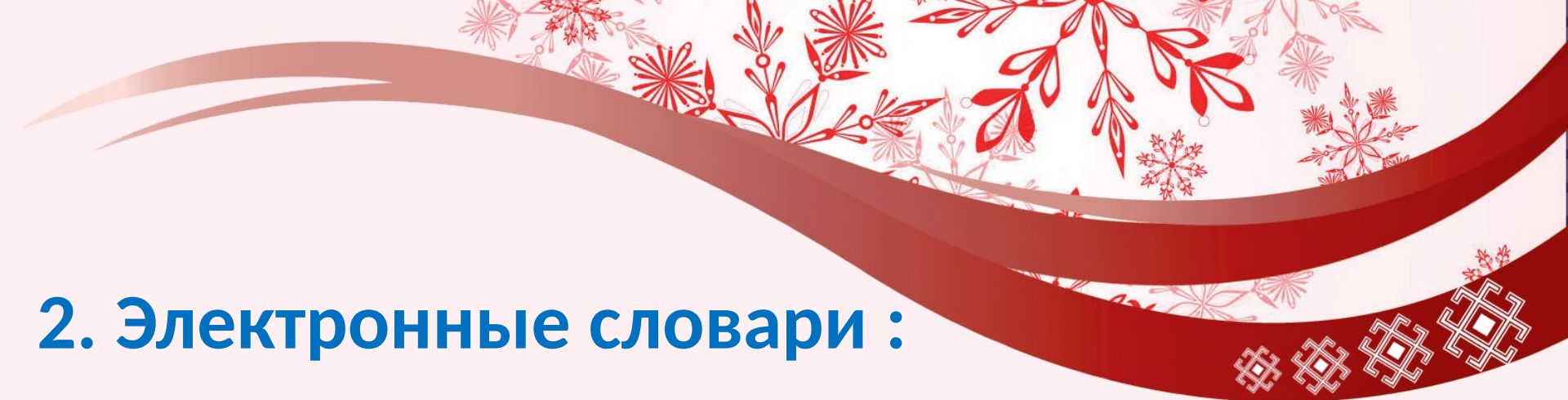
Полнотекстовый поиск ▾

https://dict.fu-lab.ru



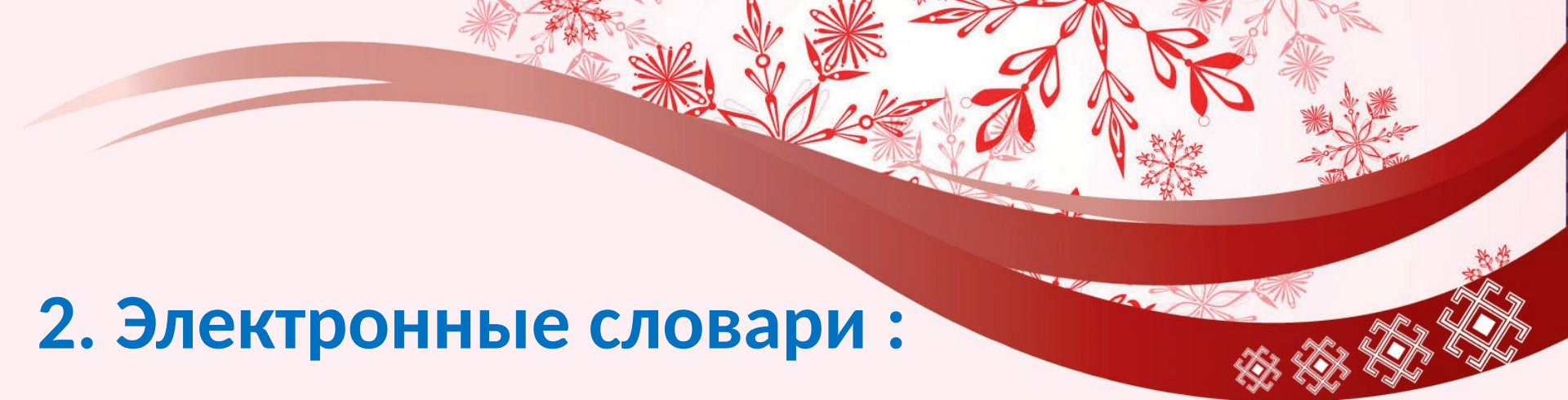
2. Электронные словари:

- «Марий йылме мутер» / «Словарь марийского языка» в 10 томах. Йошкар-Ола: Марийское книжное издательство, 1990-2005;
- А.А. Саваткова. Словарь горномарийского языка. — Йошкар-Ола. Марийское книжное издательство. 2008. — 404 с.;
- «Мутвундынам пойдарена: у мут-влак мутер». — Йошкар-Ола: Марий книга савыктыш, 2001. — 176 с.;
- Пособие по делопроизводству на марийском языке. Кагазвиктыш сомылым марла шуктышо-влаклан полшык / Составитель И.Г. Иванов. — Йошкар-Ола: Министерство культуры, печати и по делам национальностей РМЭ, 2007. — 96 с.;
- В.М. Васильев, З.В. Учаев «Марийско-русский словарь». Йошкар-Ола: Марийское книжное издательство, 2003. — 296 с.;



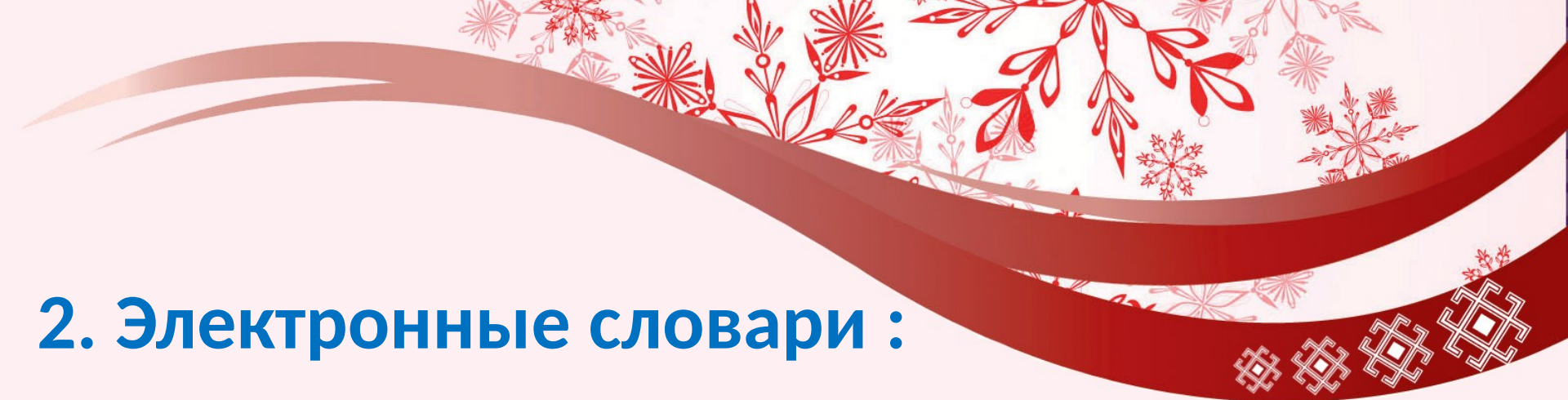
2. Электронные словари :

- Учаев З.В., Захарова В.С. Русско-марийский словарь. Йошкар-Ола: Марийское книжное издательство, 1999. — 480 с.;
- И.Г. Иванов, Е.Н. Мустаев. Марий синоним мутер. Йошкар-Ола, 1975;
- К.Г. Юадаров. Русско-марийский словарь для учащихся горномарийских средних школ. Йошкар-Ола. Марийское книжное издательство. 1979 г.;
- С.Д. Дмитриев. Русско-марийский словарь для школьников. Йошкар-Ола. Марийское книжное издательство. 2013.;
- И.Г. Иванов. Мутвундынам пойдарена: шуэн вашлиялтше мут-влак. Обогащаем лексику. Словарь редких слов марийского языка.
- Йошкар-Ола: Марий книга савыктыш, 2015. — 404 с.



2. Электронные словари :

- Е.Н. Мустаев. Словарь синонимов марийского языка. Марий синоним мутер. — Йошкар-Ола: Марийское книжное издательство, 2000. — 544 с.;
- Топонимика Республики Марий Эл: историко-этимологический анализ / О.П. Воронцова, И.С. Галкин. — Йошкар-Ола : Издательство Марийского полиграфкомбината, 2002. — 424 с.;
- Словарь сравнений марийского языка / Абукаева Любовь Алексеевна, Чемышев Андрей Валерьевич, Митрускова Анжелика Ананьевна. — Йошкар-Ола: Издательский дом «Марийское книжное издательство», 2019. — 493 с.;



2. Электронные словари :

— Русско-марийский словарь: [в 2-х томах]. Том I. А–О / авт.-сост.: Е.М. Андрианова, В.Г. Гаврилова, Э.В. Гусева и др. – Йошкар-Ола: МарНИИЯЛИ, 2019. — 498 с. Том II. П–Я / авт.-сост.: Л.И. Барцева, В.Г. Гаврилова, М.Т. Ипакова и др. – Йошкар-Ола: МарНИИЯЛИ, 2019. — 500 с.

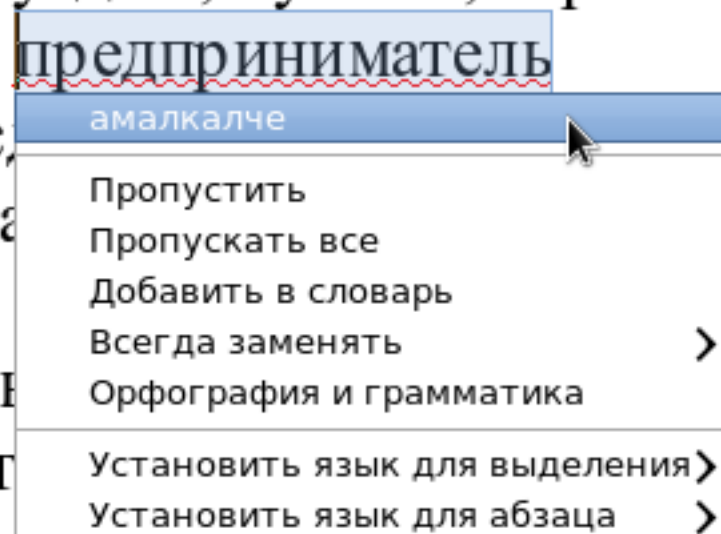
— Словарь эпитетов марийского языка / Л.А. Абукаева, А.В. Чемышев, А.А. Митрускова; МарНИИЯЛИ. — Йошкар-Ола, 2022. — 408 с.

3. Спелл-чекеры:

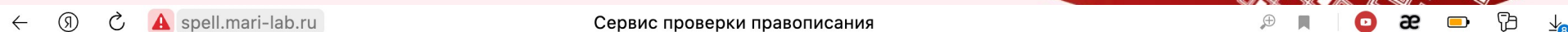
— LibreOffice:

<http://mari-lab.ru/>

Павлов. – Кургым пурташ, пуэдаш, йүкташ, терысым
Шöрым Кужэнгер районысо предприниматель
ик кечыште 80 литр марте с
нат куандара, ончыкылан үша
ялысе пörъен-влак Павловме
е, вет кызытеш рулон-влак т
ш палемденыт. Кудывечыштак чапле мончам нолтат,
ивым ыштыме, моторын сöрастарыме.



3. Спелл-чекеры:



Сервис проверки правописания

Язык: Марийский (mhr)

Окса полыш оласе администраций деч толеш. Кызыт полатыште окнам да леведышым вашталтат. Полат көргымат шот дене олмыктат. Тушто икмыняр пӧлемым посна тематике дене лабораторийыш савырат. Умбакеже нунын негызеш мастер-классым, музык курсым, театрал ден медийщик-влакын погынымашыштым эртарат. Олмыктымо паша деч вара полатыште шуко мероприятий оксам тӱлыде лиеш.

<http://spell.mari-lab.ru/>

>>>

Окса полыш оласе администраций деч толеш. Кызыт полатыште окнам да леведышым вашталтат. Полат көргымат шот дене олмыктат. Тушто икмыняр пӧлемым посна тематике дене лабораторийыш савырат. Умбакеже нунын негызеш мастер-классым, музык курсым, театрал ден **медийщик-влакын** погынымашыштым эртарат. Олмыктымо

4. Машинные переводчики :

The screenshot displays the Yandex Translate web application. At the top, the browser's address bar shows the URL `translate.yandex.ru` and the page title "Наш путь извилист, но перспект... – Яндекс Переводчик – словарь...". The Yandex logo and navigation links (Текст, Сайты, Документы, Картинки, Для бизнеса) are visible. The main interface features two text input areas. The left area, labeled "РУССКИЙ", contains the text "Наш путь извилист, но перспективы светлые." and a close button (X). The right area, labeled "МАРИЙСКИЙ", contains the translated text "Мемнан корно кадыр-кудыр, но волгыдо ончыкылыкшо." and a copy icon. Below the input areas are audio playback controls, a character count "42 / 10000", and a keyboard icon. At the bottom right of the interface, there are links for "Перевести в Google", a bookmark icon, a share icon, and thumbs up/down buttons for feedback. A small profile picture of a man with sunglasses is located in the top right corner.

сообщить об ошибке

— Яндекс.Переводчик;

4. Машинные переводчики :

translate.chv.su Атӑл Тӑлмач - переводчик на нейронном алгоритме с чувашского на русский и с русског...

Атӑл Тӑлмач ^{beta}

С мари́йского ⇌ На русский

Кузе илыштат, йолташем?

23 / 10000

Как поживаешь, друг мой?

— Атӑл Тӑлмач.

4. Машинные переводчики :

rus_11.txt Plain Text

154814 Вдруг барашек ее съел?»
154815 И вы увидите: все станет по-другому...
154816 И никогда ни один взрослый не поймет, как это важно!
154817 Это, по-моему, самое красивое и самое печальное
- место на свете.
154818 Этот же уголок пустыни нарисован и на предыдущей
- странице, но я нарисовал еще раз, чтобы вы лучше
- его разглядели.
154819 Здесь Маленький принц впервые появился на Земле, а
- потом исчез.
154820 Всмотритесь внимательней, чтобы непременно узнать
- это место, если когда-нибудь вы попадете в Африку, в
- пустыню.
154821 Если вам случится тут проезжать, заклиная вас, не
- спешите, помедлите немного под этой звездой!
154822 И если к вам подойдет маленький мальчик с золотыми
- волосами, если он будет звонко смеяться и ничего не
- ответит на ваши вопросы, вы, уж конечно,
- догадаетесь, кто он такой.
154823 Тогда — очень прошу вас! — не забудьте утешить меня
- в моей печали, скорей напишите мне, что он
- вернулся...

mhr_11.txt Plain Text

154814 Ала тага пача тудым кочкын колтен?»
154815 Да тунам те ужда: чылажат вес семын конча...
154816 Нигунамат ик кугыенат умылен ок сене, могай тиде
- кӱлешан наста!
154817 Тиде, шонымаштем, тӱняштыжат эн сӱрал да эн ойган вер.
154818 Пустыньын тиде лукшымак ончычсо лаштыкыштат
- сӱретлыме, но мый, те сайракын ончал кертса манын,
- тудым эше ик гана сӱретлышым.
154819 Изи принц Мландышке икымше гана лач тыште волен, а
- вара йомын.
154820 Африкысе пустыньыш иктаж гана логалыда гын, тиде
- верым палашак манын, сӱретым тӱслен, тӱткын ончалза.
154821 Ала тыланда ты вершӱрым эртен кудалаш перна, тунам,
- тендам сӱрвален йодам, тиде шӱдыр йымалан изишлан
- чарналтыза!
154822 Тыгодым тендан деке шӱртньӱ ӱпан изи рвезе лишемеш,
- йонгын-йонгын воштылаш тӱналеш да йодышланда
- нимомат ок вашеште гын те мутат уке, вик толдаеда, кӱ
- тиде тугай.
154823 Тунам — мый тендам моткоч йодам! — мыйым шӱлыкем
- деч лыпландараш ида мондо, мылам вашкерак серен
- колтыза: тудо пӱртылын...

— 365 тыс. предложений.

5. Синтезаторы речи:

Работа над созданием синтезатора марийской речи

Для создания аудиокорпуса построили студию звукозаписи с хорошей звукоизоляцией, звукопоглощением и звукорассеиванием.

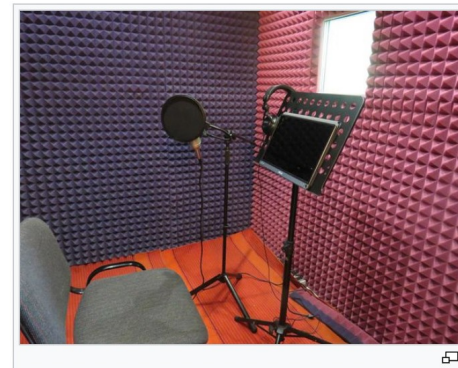
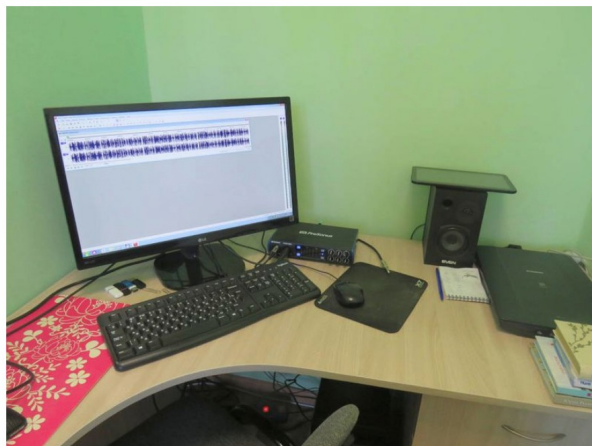
Студию комплектовали следующим необходимым оборудованием:

- студийные микрофоны;
- охватывающие наушники (головные телефоны) закрытого типа;
- выносной usb-монитор;
- предусилитель для наушников;
- микшерный пульт;
- профессиональный аудиоинтерфейс;
- микрофон для связи с диктором;
- компьютер или ноутбук с звуковым редактором;
- стойки, соединительные кабели и т.д.

Запись и обработку аудиофайлов проводим в программе Sound Forge.

Для пополнения аудиокорпуса (<https://cloud.mail.ru/public/2f76/4UHPjGjs7/>), в основном, озвучиваем произведения марийских авторов или переводные рассказы и повести, например, произведения Антона Павловича Чехова, переведённые на марийский язык в 1950 годах. Записанное в

аудиоредакторе произведение обрабатываем следующим образом: нормализуем, компрессируем. Далее, для того чтобы данные аудиофайлы являлись не только частью для аудиокорпуса, но и имели это основное значение — аудиотекст на марийском языке. Для максимального охвата потенциальной аудитории

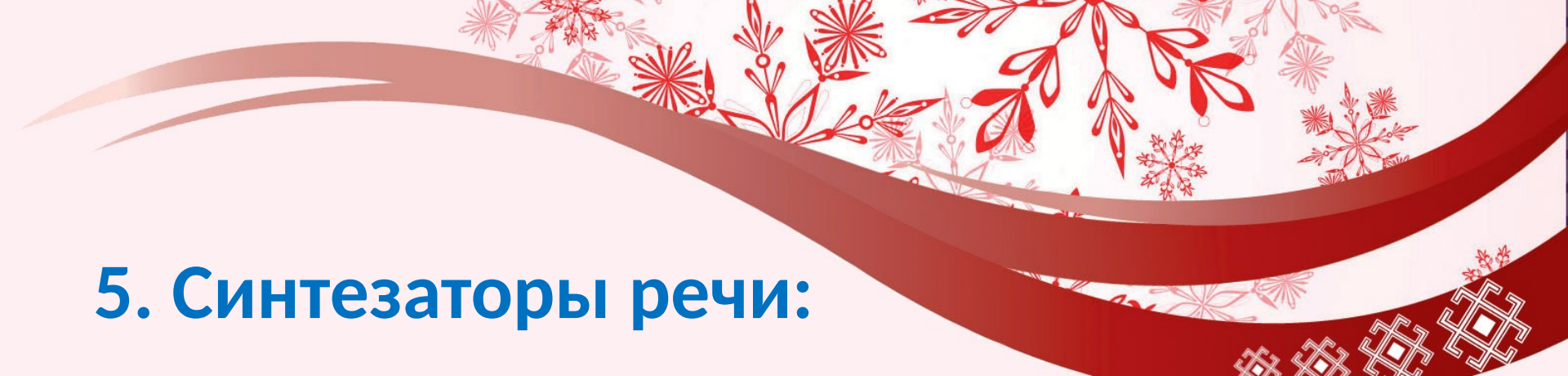


Заглавная страница
Свежие правки
Случайная статья
Справка

Инструменты
Ссылки сюда
Связанные правки
Загрузить файл
Спецстраницы
Версия для печати
Постоянная ссылка
Сведения о странице

— <http://mari-lab.ru/>

5. Синтезаторы речи:





← ↻ ↺ ⚠ mari-lab.ru

Работа над созданием синтезатора марийской речи — Mari-Lab

Андрей Обсуждение Настройки Список наблюдения Вклад Выйти

Статья Обсуждение

Читать Править История Ещё

Искать в Mari-Lab

Работа над созданием синтезатора марийской речи

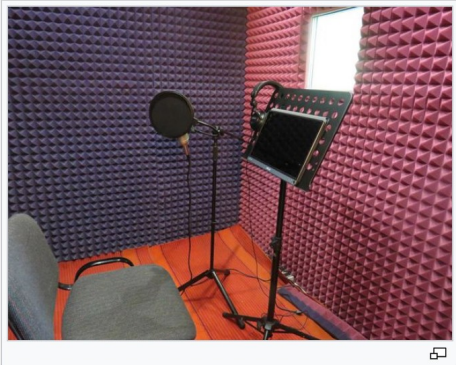
Для создания аудиокорпуса построили студию звукозаписи с хорошей звукоизоляцией, звукопоглощением и звукорассеиванием.

Студию комплектовали следующим необходимым оборудованием:

- студийные микрофоны;
- охватывающие наушники (головные телефоны) закрытого типа;
- выносной usb-монитор;
- предусилитель для наушников;
- микшерный пульт;
- профессиональный аудиоинтерфейс;
- микрофон для связи с диктором;
- компьютер или ноутбук с звуковым редактором;
- стойки, соединительные кабели и т.д.

Запись и обработку аудиофайлов проводим в программе Sound Forge.

Для пополнения аудиокорпуса (<https://cloud.mail.ru/public/2f76/4UHPjGjs7/>), в основном, озвучиваем произведения марийских авторов или переводные рассказы и повести, например, произведения Антона Павловича Чехова, переведённые на марийский язык в 1950 годах. Записанное в аудиоредакторе произведение обрабатываем следующим образом: нормализуем, компрессируем. Далее, для того чтобы лангские аудиофайлы являлись не только частью аудиокорпуса, но и мультимедийное содержание — аудиокниги на марийском языке для максимальной эффективности восприятия аудиторией.



Заглавная страница

Свежие правки

Случайная статья

Справка

Инструменты

Ссылки сюда

Связанные правки

Загрузить файл

Спецстраницы

Версия для печати

Постоянная ссылка

Сведения о странице

— обученная модель на Google/Tacotron.

6. Распознавание речи:

Common Voice

НАДЫРЛАШ ДАТАСЕТ-ВЛАК ЙЫЛМЕ-ВЛАК МЕМНАН НЕРГЕН

★ 208 ОТЗЫВОВ

1030 38113

АНДРЕЙ ЧЕМЫШЕВ

Кутыраш

Шке йүкетым
надырле

Колышташ

Мыланна йүк-влакым
чынлаш полшо

— Mozilla Common Voice.

6. Распознавание речи:

← ⓘ ↻ commonvoice.mozilla.org Common Voice 🔍 🗖️ 🎧 208 отзывов 📖 🌐 ⚙️

← Кутыраш Колышташ

🔗 НАДЫРЛЫМЕ ВИСЫКЫМ УМЫЛЕН НАЛАШ

Пытартыш!

«Турий кочкишым сайын
ямдыла, нимом от ман.

▶ ↺ ↻ 1

▶ ↺ ↻ 2

▶ ↺ ↻ 3

▶ ↺ ↻ 4

ВОЗАШ ТҰНАЛАШ 5

🎤

6. Распознавание речи:

← ⓘ ↻ commonvoice.mozilla.org Common Voice 🔍 🗲 🎧 ★ 208 отзывов 📖 📺 ⌂ ⌵

← Кутыраш Колышташ ↗ НАДЫРЛЫМЕ ВИСЫКЫМ УМЫЛЕН НАЛАШ

Темдал ▶ нуно ойлончым чын йўкаңденыт?

Тудо пушенге-влак коклаште
шинча, тудлан кошташат ок
кўл.

1 2 3 4 5

👍 ТУГЕ ▶ 🗨 УКЕ



Проблемы:

— нехватка ресурсов, необходимо собрать и выровнять 1 миллион предложений в параллельном корпусе, на сегодняшний день собрано 365 тысяч предложений.

— ...



Перспективы:

- ГОЛОСОВЫЕ ПОМОЩНИКИ;
- система автоматического перевода аудио- и видеоконтента с марийского и на марийский язык;
- ...

