Поиск по архивам

Поиск по архивам

Проект Яндекса «Поиск по архивам»

Проект, помогающий находить упоминания о людях, населённых пунктах, событиях и явлениях в архивных документах при помощи собственной нейросетевой технологии компании

YOY, АПРЕЛЬ 2025 → АПРЕЛЬ 2026

16 → 24
региона России присутствуют на сервисе

16 → 22 млн
страниц архивных документов

4 → 6 млн
страниц периодических и справочных изданий

87 → 90%
точность распознавания архивных документов*

33 → 80%
точность распознавания скорописи

87 → 96%
точность распознавания газет и периодических изданий

500 тыс.
человек — средняя ежемесячная аудитория

100 лет
истории советского и российского спорта доступно на сервисе

ЯНВ'2023 — ЗАПУСК СЕРВИСА

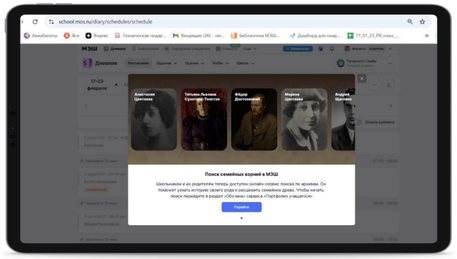
2 из 18
субъекты Центрального федерального округа

41%
архивных документов на сервисе приходится на Москву и МО

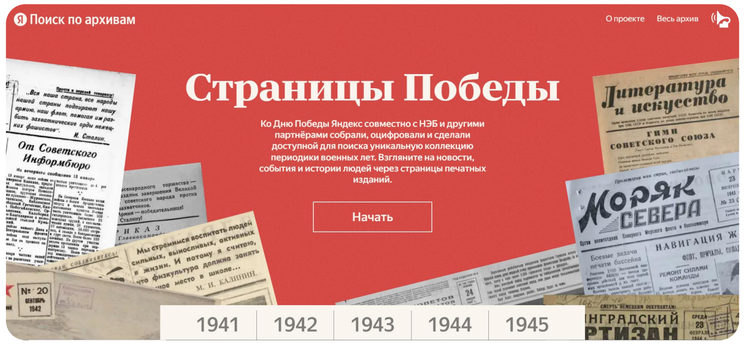
5 из 14
субъекты ПФО

4 из 10
субъекты СФО

4 из 11
субъекты ДВФО



Сервис интегрирован в Московскую электронную школу (МЭШ)



Сделали главный спецпроект Яндекса к празднику 9 мая в 2025 году под названием «Страницы Победы» совместно с партнёрами.

900 000
человек посетило страницу проекта в первый день

* Лучший показатель на рынке

Уникальность проекта

Мы единственный сервис в мире, обладающий таким функционалом

Архивные АИС

- Отсутствует поиск по тексту рукописных документов
- Отсутствует/присутствует в малом количестве индексация дел
- Технические ограничения

Частные проекты «Генотек.Архивы» и «hryc.by»

- Хуже качество распознавания
- Спорные моменты по юридической части

Проекты ручной индексации «Великие описи» и Familio

- Несопоставимый масштаб из-за опоры на ручной труд

Международный проект «Familysearch.org»

- Плохая организация документов на сервисе
- Плохое качество микрофильмированных материалов
- Зачастую непонятное описание материалов, сложность в поиске
- Нет доступа из России

Глобальная цель проекта

Проект «Поиск по архивам» делает искусственный интеллект более **человечным**, **точным** и **отечественным**

За счёт изучения архивных документов и периодических изданий ИИ:

- Получает доступ к первоисточникам и проверенным данным
- Снижает «галлюцинации» при ответах об истории России
- «Мыслит как русский человек»

- Лучше понимает российский контекст и реалии
- Узнает о специфических событиях, местах и персоналиях
- Учитывает культурные особенности разных эпох истории страны

Наши партнёры

Мы работаем не только
с региональными архивами,
но и крупными библиотеками

РОССИЙСКАЯ
ГОСУДАРСТВЕННАЯ
БИБЛИОТЕКА

РОССИЙСКАЯ
НАЦИОНАЛЬНАЯ
БИБЛИОТЕКА

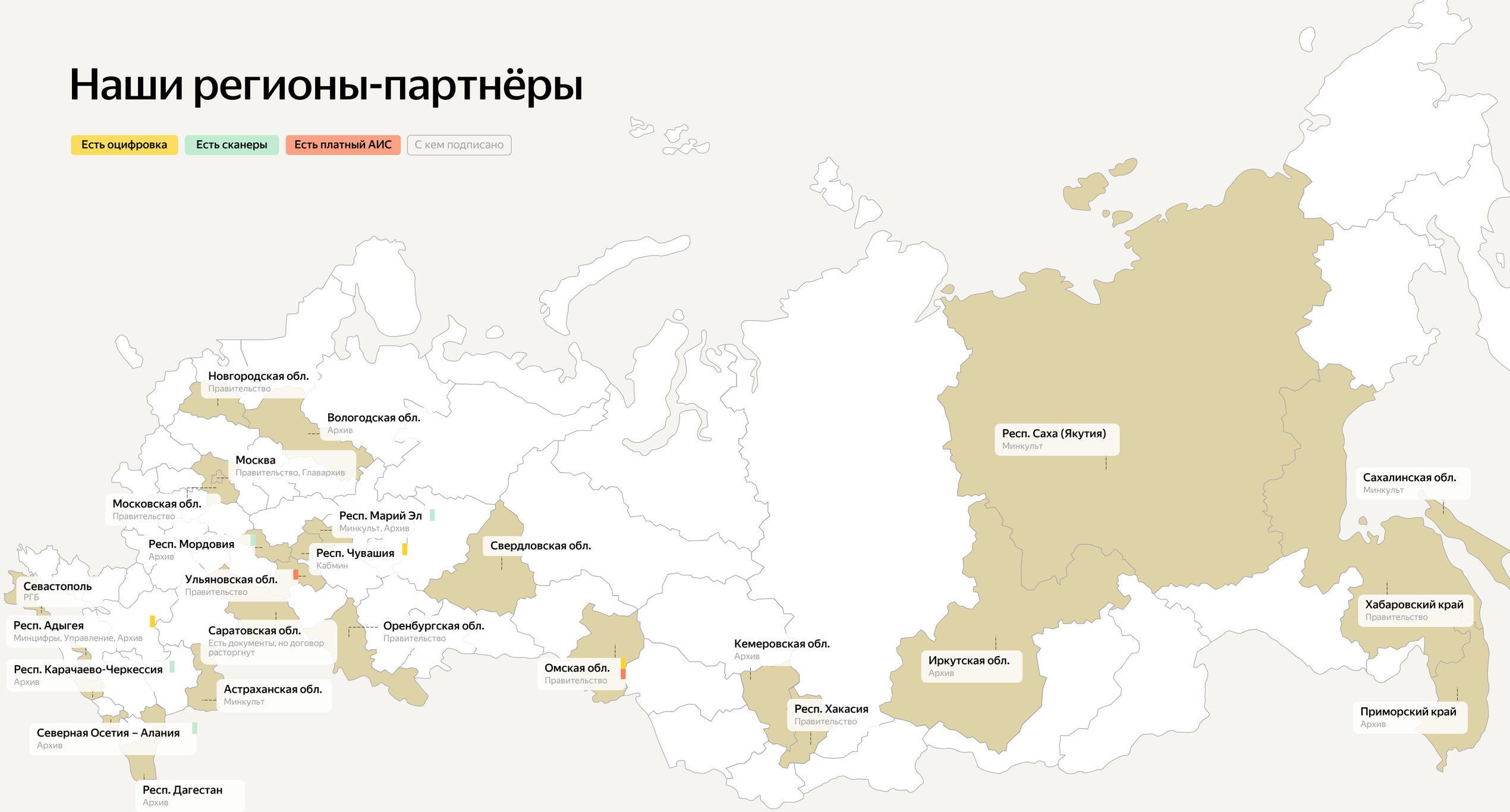
ПРЕЗИДЕНТСКАЯ
БИБЛИОТЕКА

РУНИВЕРС

Библиотека
им. Н. А. Некрасова

Наши регионы-партнёры

Есть оцифровка | Есть сканеры | Есть платный АИС | С кем подписано



Владимир Путин объявил 2026 годом единства народов

В России проживают представители 194 национальностей, которые говорят на более чем 300 языках и диалектах, из них 150 относятся к языкам народов России.

По данным регулярно проводимых исследований использование этих языков стремительно сокращается

Проект «Языки народов России»

Поиск

2023
запуск сервиса

390 000
уникальных пользователей в месяц в Переводчике

540 000
уникальных пользователей в месяц в Поиске

9,8 млн
человек — носители поддержанных языков

На данный момент доступен перевод для 18 языков

Башкирский · Татарский · Чувашский · Марийский · Горномарийский · Удмуртский · Якутский · Осетинский · Тувинский · Бурятский · Коми · Мокшанский · Эрзянский · Кабардино-черкесский · Карачаево-балкарский · Мансийский · Абазинский · Ногайский

Синтез и распознавание для 5 языков

Башкирский · Татарский · Чувашский · Удмуртский · Марийский

Цель на 3 года

Добавить 20 языков народов России в Переводчик, а также синтез речи и распознавание для 10 языков

Главы регионов положительно отзываются о проекте и ценят его вклад в развитие и сохранение культуры

Планы 2026

2026 — год единства народов России

Добавим перевод

- Чеченский
- Лезгинский
- Адыгейский
- Ингушский
- Алтайский
- Хакасский
- Карельский

Добавим синтез и распознавание

- Якутский
- Кабардино-черкесский
- Бурятский
- Тувинский

Доступные языки на данный момент

Язык	Дата релиза	Переводчик	Синтез + распознавание	Клавиатура	Фотоперевод
Татарский	8 июня 2015	+	+	+	+
Башкирский	2 сент 2015	+	+	+	+
Удмуртский	6 апр 2016	+	+	+	+
Марийский	6 сент 2016	+	+	+	+
Горномарийский	6 сент 2016	+	−	+	−
Чувашский	11 фев 2020	+	+	+	+
Якутский	27 апр 2020	+	−	+	+
Осетинский	2 июля 2024	+	−	+	−
Коми	10 фев 2025	+	−	+	−
Тувинский	26 мар 2025	+	−	+	−
Эрзянский	3 июля 2025	+	−	+	−
Мокшанский	3 июля 2025	+	−	+	−
Карачаево-балкарский	24 сент 2025	+	−	+	−
Кабардино-черкесский	24 сент 2025	+	−	+	−
Бурятский	24 окт 2025	+	−	+	−
Абазинский	10 дек 2025	+	−	+	−
Ногайский	10 дек 2025	+	−	+	−
Мансийский	10 дек 2025	+	−	+	−

Количество уникальных пользователей, мес.

Язык	Перевод, тыс.	Тематический блок, тыс.	Считают родным, тыс.
Татарский язык	168,7	337,3	4 073
Башкирский язык	59	72,1	1 320
Чувашский язык	38	40,1	800
Якутский язык	24	16,6	479
Кабардино-черкесский	20	21,9	609
Карачаево-балкарский	16,3	17,7	343
Удмуртский язык	11,3	14,6	272
Марийский язык	11,1	16,3	318
Осетинский язык	10,7	12,4	457
Горномарийский язык	7,5	1,7	36,8
Тувинский язык	6,9	8,4	294
Эрзянский	6,2	7,8	230
Мокшанский	5,9	7,1	115
Коми язык	4,8	5,8	100
Бурятский	15,8	28,6	393
Абазинский	18,4	48,7	37
Ногайский	28,6	29,7	87,1
Мансийский	3,1	8,1	0,229
Итого	**456,3**	**694,9**	**9 966,10**

Качество перевода

Мы лучше Google
и сторонних существующих
переводчиков

BLEU

Bilingual Evaluation Understudy

Это метрика для оценки качества текста, который был машинно переведён с одного естественного языка на другой

Статистика по BLEU	Яндекс	Google	Claude Sonnet
Татарский — Русский	23,4	24,8	23,5
Русский — Татарский	17,3	13	16,1
Марийский — Русский	43,3	32	39,2
Русский — Марийский	29,5	15,4	17,5
Осетинский — Русский	20,6	17,6	22,5
Русский — Осетинский	10	8,1	7,6
Бурятский — Русский	42,8	29,9	30,9
Русский — Бурятский	18,5	8,8	7,4
Кабардино-черкесский — Русский	26,2	12,2	27,5
Русский — Кабардино-черкесский	11,3	1,7	9,4

Google Переводчик — онлайн-сервис от компании Google, который предоставляет возможность переводить тексты, веб-страницы, а также документы и голосовые разговоры на разные языки автоматически.
Claude Sonnet — семейство больших языковых моделей (LLM) искусственного интеллекта (ИИ), разработанных компанией Anthropic.
Данные приведены для версии 3.5, в текущий момент компания использует версии 4 и 4.5 с менее продвинутым переводом.

Партнёры проекта

ФАДН РОССИИ
Федеральное агентство по делам национальностей

ДОМ НАРОДОВ РОССИИ

Мы работаем с языковыми институтами, научными центрами и языковыми активистами

1. Карачаево-Черкесский институт гуманитарных исследований имени Х. Х. Хапсирокова
2. Республиканский центр по поддержке изучения национальных языков и иных предметов этнокультурной направленности «Бэлиг»
3. Карельская региональная общественная организация «Общество вепсской культуры»
4. Кабардино-Балкарская региональная общественная организация по содействию развитию адыгской молодежи «Черкесский Ренессанс»
5. ГАУ РК «Дом дружбы народов Республики Коми»
6. Обско-угорский институт прикладных исследований и разработок совместно с Югорским научно–исследовательским институтом информационных технологий
7. Мордовский государственный университет им. Н. П. Огарева
8. Национальная библиотека Республики Саха (Якутия)
9. ГКУ «Институт чеченского языка»
10. ГБНИиОУ «Тувинский институт гуманитарных и прикладных социально-экономических исследований при Правительстве Республики Тыва» совместно с РОО «Совет молодых учёных и специалистов Республики Тыва»
11. Удмуртский федеральный исследовательский центр Уральского отделения РАН
12. Академия наук Республики Татарстан
13. Северо-Осетинский государственный университет им. К. Л. Хетагурова