



**ФЕДЕРАЛЬНАЯ СЛУЖБА  
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ  
ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ БАЗЫ ДАННЫХ, ОХРАНЯЕМОЙ  
АВТОРСКИМИ ПРАВАМИ**

Номер регистрации (свидетельства):  
2023624102

Дата регистрации: 22.11.2023

Номер и дата поступления заявки:  
2023623762 08.11.2023

Дата публикации и номер бюллетеня:  
22.11.2023 Бюл. № 12

Контактные реквизиты:  
нет

Автор(ы):

Кипяткова Ирина Сергеевна (RU),  
Родионова Александра Павловна (RU),  
Кагиров Ильдар Амирович (RU),  
Крижановский Андрей Анатольевич (RU)

Правообладатель(и):

Федеральное государственное бюджетное  
учреждение науки "Санкт-Петербургский  
Федеральный исследовательский центр  
Российской академии наук" (RU)

Название базы данных:

**База данных аннотаций речевых записей на карельском языке (AnKaS – Database of Annotations of Karelian Speech Recordings)**

**Реферат:**

База данных (БД) представляет собой набор аннотаций записей карельской речи (ливвиковское наречие) из 13 радиопередач программы «Kodirandaine» («Родной берег»). Оригинальные аудиофайлы находятся в открытом доступе на сайте ГТРК «Карелия»:

<https://tv-karelia.ru/kodirandaine-rodnoy-bereg/>. БД составлена в формате JSON. Аннотация записей речи каждого диктора содержится в отдельном .json файле и включает в себя следующую информацию: номер фразы; интернет-ссылку на аудиозапись; время начала фразы; время конца фразы; текстовую расшифровку; текстовую расшифровку с тегами, отмечающими другой языковой код (русский). Всего в БД содержатся аннотации 4385 фраз. БД предназначена для обучения и тестирования системы автоматического распознавания карельской речи. БД разработана при финансовой поддержке Российского научного фонда, проект № 22-21-00843. Тип ЭВМ: IBM PC-совмест. ПК; ОС: Windows, Linux.

**Вид и версия системы управления базой данных:**

Специализированная СУБД СПб ФИЦ РАН

**Объем базы данных:**

1,6 МБ