



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ БАЗЫ ДАННЫХ, ОХРАНЯЕМОЙ АВТОРСКИМИ ПРАВАМИ

Номер регистрации (свидетельства):
2018620585

Дата регистрации: 18.04.2018

Номер и дата поступления заявки:
2017621488 26.10.2017

Дата публикации и номер бюллетеня:
18.04.2018 Бюл. № 4

Контактные реквизиты:
Нет

Автор(ы):

Конев Антон Александрович (RU),
Шумская Ольга Олеговна (RU)

Правообладатель(и):

Федеральное государственное бюджетное
образовательное учреждение высшего
образования «Томский государственный
университет систем управления и
радиоэлектроники» (ТУСУР) (RU)

Название базы данных:

Речевой корпус для систем анализа и распознавания речи

Реферат:

База данных представляет собой комплекс из 27 таблиц, хранящих информацию об аудиофайлах: их содержании и сегментации, о дикторах. База данных предназначена для проведения исследований в области распознавания речи и идентификации дикторов. База данных позволяет осуществлять просмотр или дополнение уже существующих записей, а также добавление новых аудиофайлов и их сегментацию. При заполнении базы данных речевого корпуса подбор и расположение материалов осуществлялись по следующим требованиям: наличие дикторов разной диалектической принадлежности; наличие дикторов различных возрастов и пола; учет дикторов с дефектами речи, заболеваниями речевого тракта, акцентом; наличие сигналов с различной эмоциональной окраской; возможность пополнения созданного корпуса новыми сигналами; наличие ручной сегментации ряда речевых сигналов для контроля качества автоматической сегментации. База данных содержит: 77 речевых единиц, данные о 193 дикторах, 124 речевых сигнала общей длительностью 842 секунды, результаты ручной сегментации для 103 речевых сигналов.

Вид и версия системы управления базой данных: SQL Server 2012

Объем базы данных: 100 032 Кб