

## **Настройки распознавания речи с помощью движка Vosk**

## ОГЛАВЛЕНИЕ

1	Общие сведения .....	3
2	Взаимодействие компонентов .....	4
3	Аппаратные требования.....	5
4	Порядок настройки.....	7

## 1 ОБЩИЕ СВЕДЕНИЯ

В LibraDLP имеется возможность использовать для распознавания речи движок **Vosk** – **open-source решение**, позволяющее **лучше масштабировать производительность** системы при распознавании аудиоданных.

Для использования **Vosk** не нужно приобретать лицензии – компонент не лицензируется. Это позволяет распознавать звуковые данные за **любой** период (в том числе архивные данные) из любого канала перехвата: *мессенджеры, звуковые файлы на внешних устройствах хранения (USB), звуковые файлы из папки (индексация локальных папок) и т.д.*

Интеграция инструмента в LibraDLP с поддержкой **различных языковых моделей** распознавания значительно **расширяет** возможности DLP-системы. При использовании Vosk количество потоков практически не влияет на использование оперативной памяти: для большой модели Vosk в среднем требуется 6-8 ГБ ОЗУ.

При использовании CPU один поток использует одно ядро. Например, на рабочей станции с 16 ядрами и 16 ГБ памяти можно максимально запустить 12 потоков, 3i Speech на этих же ресурсах будет работать в 1 поток. Также в сравнении с 3i Speech настройка движка Vosk значительно проще.

Для использования движка и просмотра результатов распознавания необходимо наличие следующих версий модулей LibraDLP:

- DataCenter 2.57.6.6 и выше.
- SearchServer 5.27.0.8 и выше.
- AAServer 2.3.0.20 и выше.
- AnalyticConsole 1.79.1.1 и выше.

## 2 ВЗАИМОДЕЙСТВИЕ КОМПОНЕНТОВ

Для взаимодействия компонентов системы должны быть открыты порты, представленные в Таблице 1.

Таблица 1 – Порты взаимодействия компонентов

Компонент	Протокол/Порт	Компонент	Протокол/Порт
<b>AAServer</b>			
<b>Взаимодействие с сервером баз данных MS SQL</b>			
Сервер AAServer	TCP/*	Сервер баз данных MS SQL	TCP/1433
	UDP/*		UDP/1434
<b>Взаимодействие с сервером баз данных PostgreSQL</b>			
Сервер AAServer	TCP/*	Сервер баз данных PostgreSQL	TCP/5432
<b>Отправка и получение данных с AAServer</b>			
SearchServer AnalyticConsole	TCP/*	Сервер AAServer	TCP/9070
<b>Управление AAServer</b>			
Клиент AAServer Клиент DataCenter	TCP/*	Сервер AAServer	TCP/9070

### 3 АППАРАТНЫЕ ТРЕБОВАНИЯ

В зависимости от количества пользователей, аудиоданные которых необходимо распознавать, требуются следующие характеристики сервера для использования Vosk:

#### С малой моделью Vosk:

до **10** пользователей

CPU	от 2.5 GHz (4 ядра)
RAM	от 8 GB
HDD	от 100 GB
OS	Microsoft Windows Server 2012 R2 и выше Ubuntu 22.04 Astra Linux 1.7.3

до **20** пользователей

CPU	от 2.5 GHz (4 ядра)
RAM	от 8 GB
HDD	от 100 GB
OS	Microsoft Windows Server 2012 R2 и выше Ubuntu 22.04 Astra Linux 1.7.3

до **50** пользователей

CPU	от 2.5 GHz (8 ядер)
RAM	от 8 GB
HDD	от 100 GB
OS	Microsoft Windows Server 2012 R2 и выше Ubuntu 22.04 Astra Linux 1.7.3

до **100** пользователей

CPU	от 2.5 GHz (14 ядер)
RAM	от 8 GB
HDD	от 100 GB
OS	Microsoft Windows Server 2012 R2 и выше Ubuntu 22.04 Astra Linux 1.7.3

#### С большой моделью Vosk:

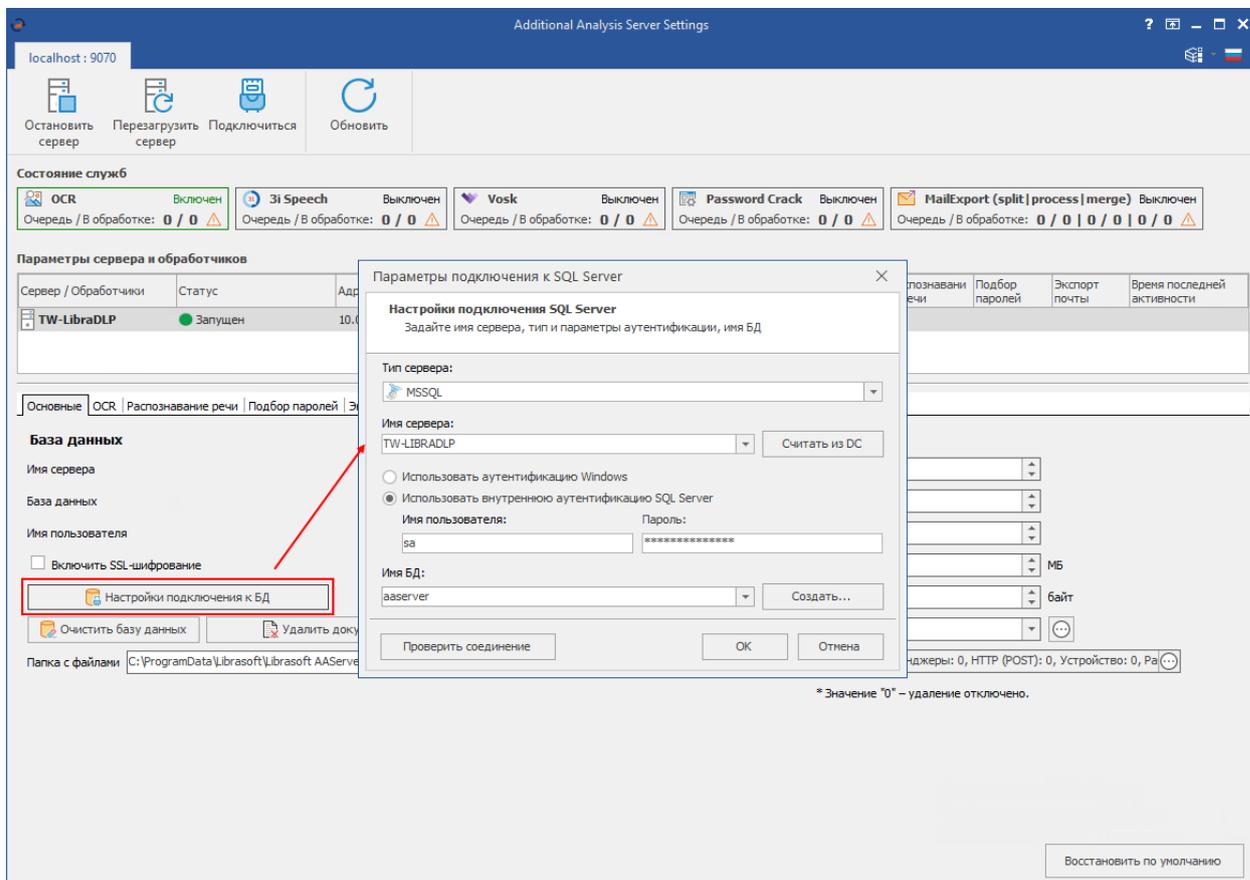
до **10** пользователей

CPU	от 2.5 GHz (4 ядра)
RAM	от 16 GB
HDD	от 100 GB

OS	Microsoft Windows Server 2012 R2 и выше
	Ubuntu 22.04
	Astra Linux 1.7.3
<b>до 20 пользователей</b>	
CPU	от 2.5 GHz (4 ядра)
RAM	от 16 GB
HDD	от 100 GB
OS	Microsoft Windows Server 2012 R2 и выше
	Ubuntu 22.04
	Astra Linux 1.7.3
<b>до 50 пользователей</b>	
CPU	от 2.5 GHz (24 ядра)
RAM	от 16 GB
HDD	от 100 GB
OS	Microsoft Windows Server 2012 R2 и выше
	Ubuntu 22.04
	Astra Linux 1.7.3
<b>до 100 пользователей</b>	
CPU	от 2.5 GHz (48 ядра)
RAM	от 24 GB
HDD	от 100 GB
OS	Microsoft Windows Server 2012 R2 и выше
	Ubuntu 22.04
	Astra Linux 1.7.3

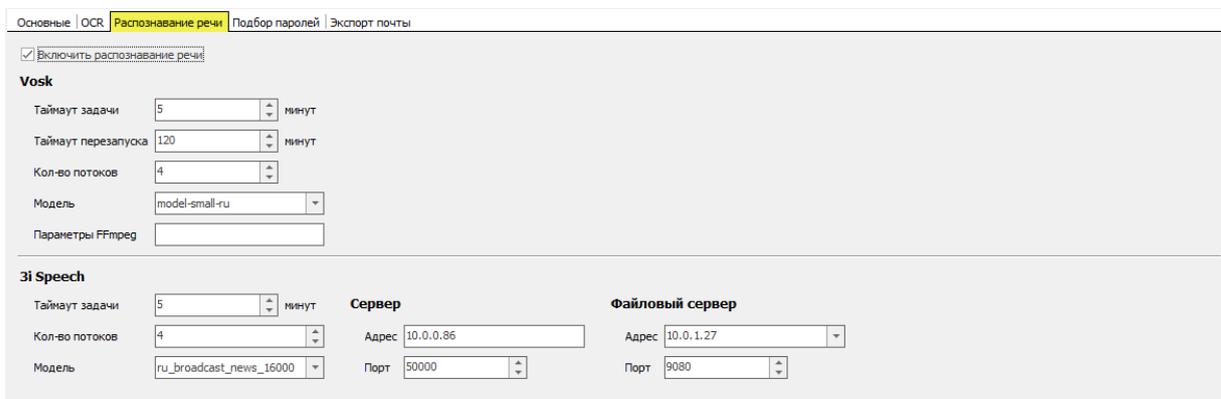
## 4 ПОРЯДОК НАСТРОЙКИ

В первую очередь в консоли **AAServer** на вкладке **Основные** проверьте наличие настроенного подключения к БД. Если подключение не настроено нажмите кнопку **Настройки подключения к БД** для создания новой базы данных, заполните параметры подключения к SQL и нажмите **Создать...** для создания БД настроек AAServer.



Для извлечения текста из аудиофайлов с помощью движка **Vosk** выполните настройки:

1. В консоли **AAServer** на вкладке **Распознавание речи** включите опцию **Включить распознавание речи**.

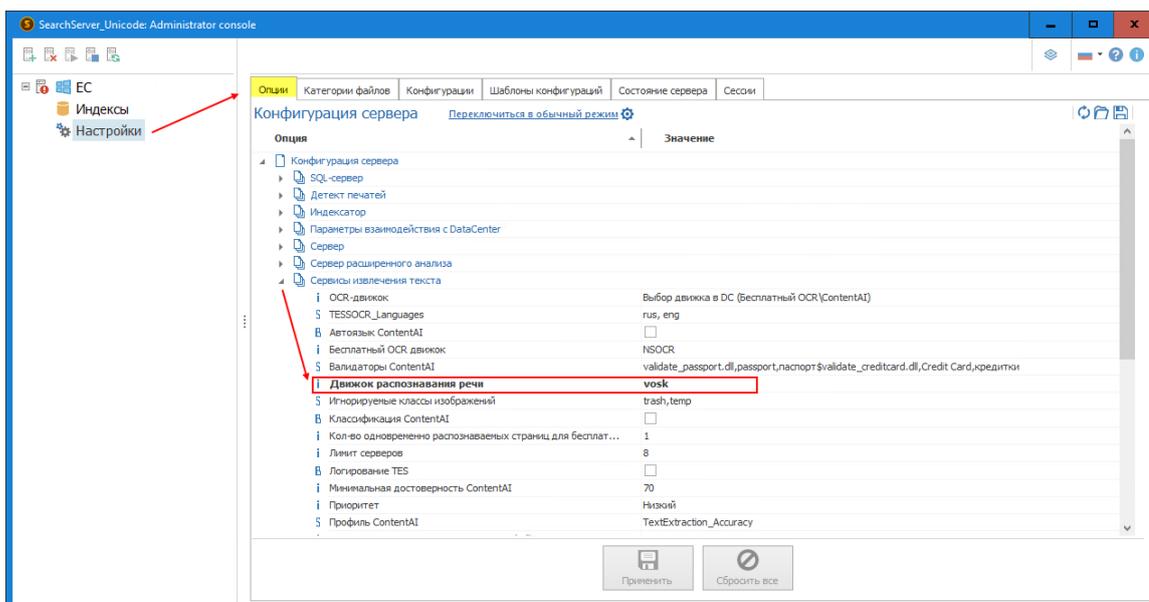


2. Задайте параметры работы движка **Vosk** (см. таблицу ниже).

Таблица 2 – Параметры движка распознавания речи Vosk

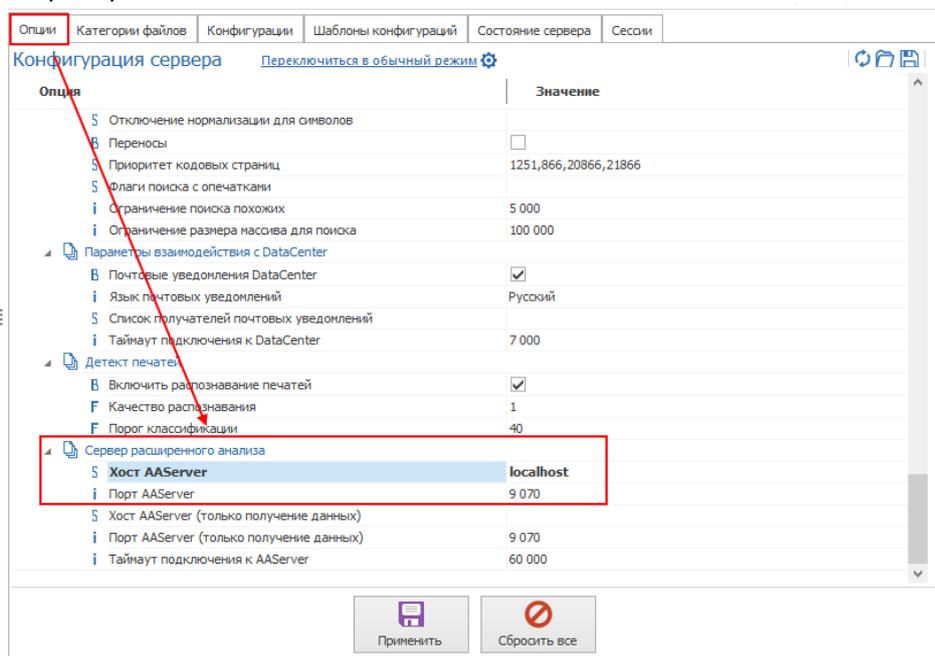
Параметр	Описание
<b>Таймаут задачи</b>	Максимальное время выполнения задачи, минут.
<b>Таймаут перезапуска</b>	Время, по истечении которого движок будет перезапущен. Значение по умолчанию - 120 минут. Доступный диапазон: 30-240 минут. Если во время распознавания наступает время перезапуска по таймауту, сервер не перезапустится, пока не завершится распознавание. Если во время работы наблюдается чрезмерное потребление оперативной памяти, рекомендуется уменьшить значение тайм-аута.
<b>Кол-во потоков</b>	<p>Количество файлов, параллельно обрабатываемых каждым воркером. Рекомендованное число потоков равно <b>0,75 * кол-во ядер</b>.</p> <div style="border: 1px dashed gray; padding: 5px;">  <p>В связи с особенностями архитектуры Vosk для распределения нагрузки между ядрами для серверов на базе <b>ОС Windows</b> существуют программные ограничения: если в системе более 30 ядер (логических процессоров), Vosk будет использовать только половину из них. Для серверов на базе ОС <b>Linux</b> ограничений нет.</p> </div>
<b>Модель</b>	<p>Набор файлов, определяющих уровень качества и язык распознавания. При установке на сервер устанавливается одна небольшая по объему модель <b>model-small-ru</b>, обеспечивающая высокую скорость обработки и базовый уровень качества.</p> <p>Из отдельных дистрибутивов можно установить <b>дополнительные русские и английские модели</b>, обеспечивающие более высокое качество распознавания. Установка дополнительных моделей осуществляется только на сервер! После этого на воркерах список моделей автоматически синхронизируется со списком на сервере. <a href="#">Ссылка</a> на скачивание большой модели Vosk.</p> <p>Также можно добавить в папку «<b>C:\Program Files\Librasoft\Librasoft AAServer\vosk\custom</b>» любую доступную модель Vosk (можно <a href="#">скачать</a> с официального сайта проекта), она будет использоваться при выборе из списка модели <b>custom</b>.</p>
<b>Параметры FFmpeg</b>	Дополнительные параметры для пакета FFmpeg для предварительной обработки распознаваемого звукового файла. По умолчанию применяются параметры <b>afftdn=tn=enable</b> и <b>loudnorm</b> .

3. Перейдите в консоль [SearchServer](#) и включите использование движка **Vosk**. Для этого перейдите в меню **Настройки** → **Экспертный режим** → вкладка **Опции** → **Сервисы извлечения текста** → параметр **Движок распознавания речи** → **Vosk**. Нажмите **Применить**.



Появится уведомление о необходимости перезагрузки службы сервера для применения новых настроек. Нажмите **ОК**.

4. В этом же разделе ниже укажите адрес AAServer, с помощью которого будет осуществляться распознавание речи: в группе параметров **Сервер расширенного анализа** укажите в поле параметра **Хост AAServer** имя или IP-адрес сервера с установленным AAServer, а в поле **Порт AAServer** – номер порта, на котором работает AAServer.



5. Появится уведомление о необходимости перезагрузки службы сервера для применения новых настроек. Нажмите **ОК**.
6. Создайте индекс выбранного источника данных, если необходимого индекса нет (если индекс имеется, то перейдите к [шагу 7](#)).  
Для этого в SearchServer выберите из контекстного меню команду **Создать индекс**.

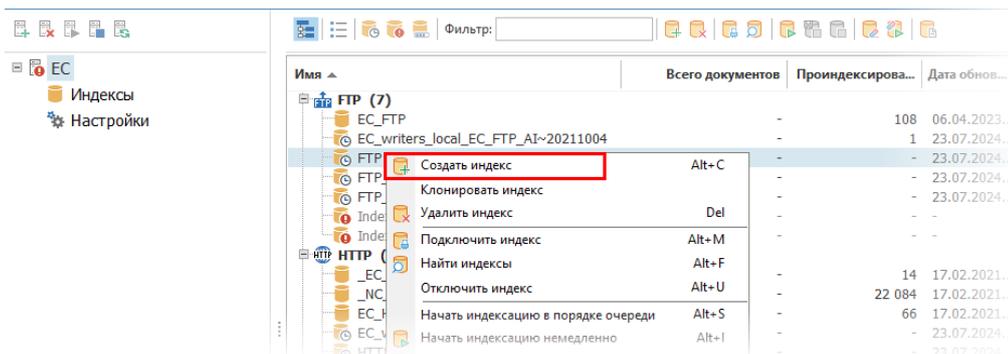
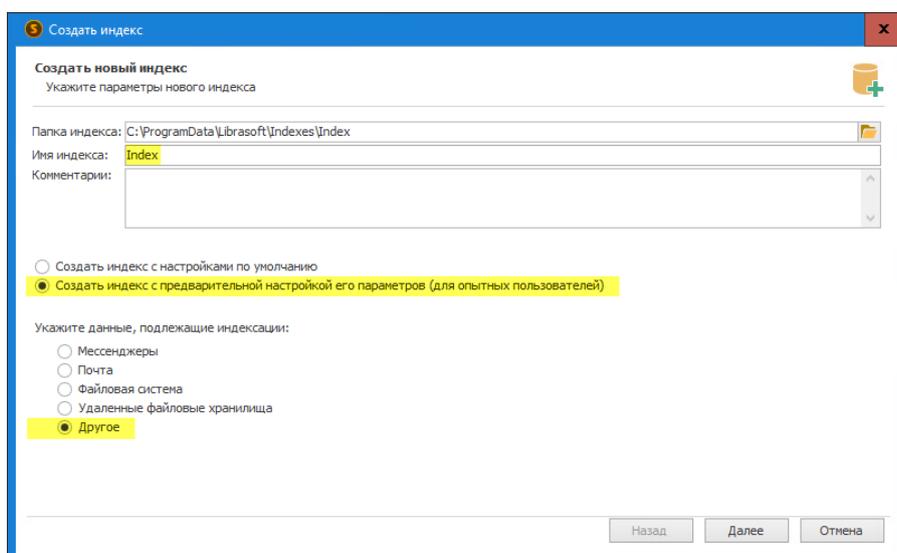


Таблица 3 – Создание индекса

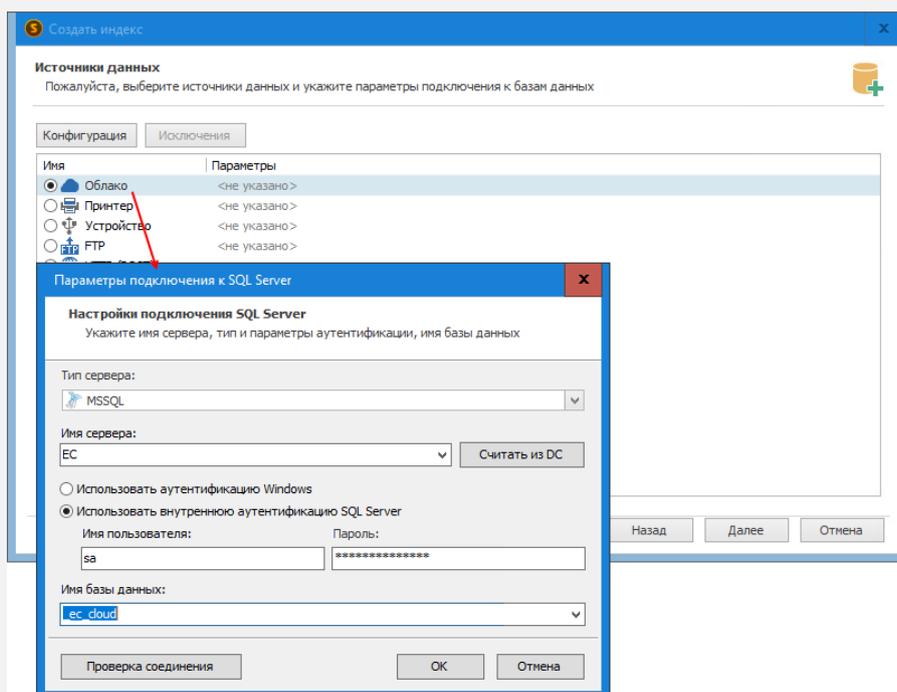
## Действие

## Шаг в мастере создания индекса

Укажите **Имя индекса**, отметьте опцию **Создать индекс с предварительной настройкой его параметров** и выберите тип источника данных **Другое**.

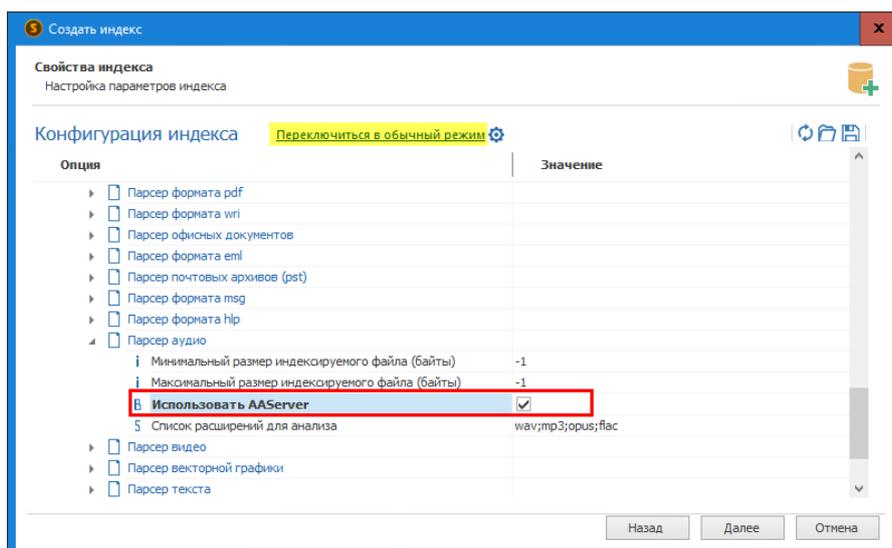


Отметьте источник данных и нажмите кнопку **Конфигурация**. Выберите базу данных, которая будет индексироваться.

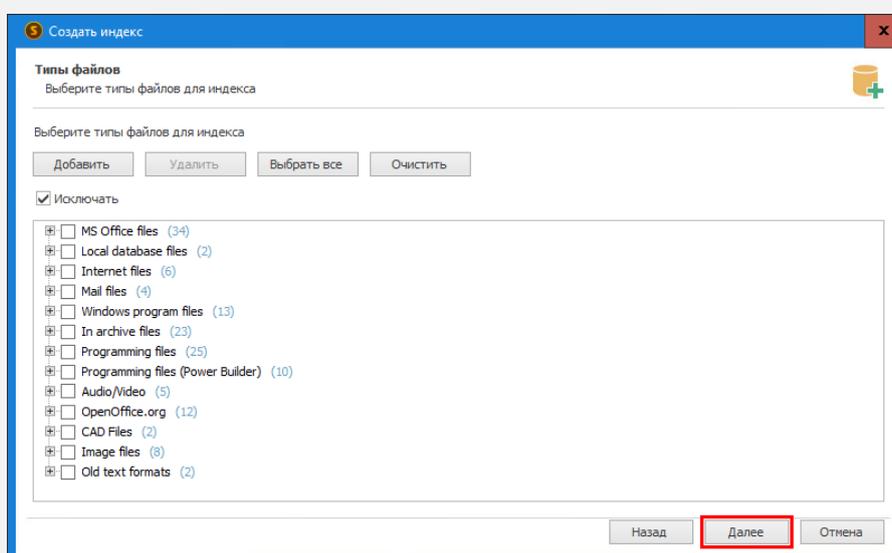


Далее в свойствах индекса переключитесь в **экспертный режим** и включите в группе **Парсеры** → **Парсер аудио** опцию **Использовать AAServer**.

При необходимости в параметре **Список разрешений для анализа** выберите типы файлов которые будут распознаваться.

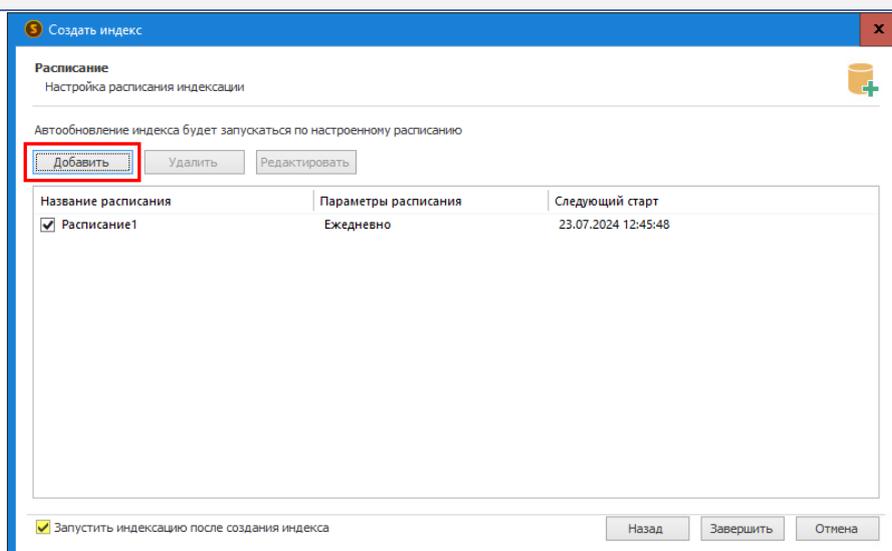


Нажмите **Далее**.

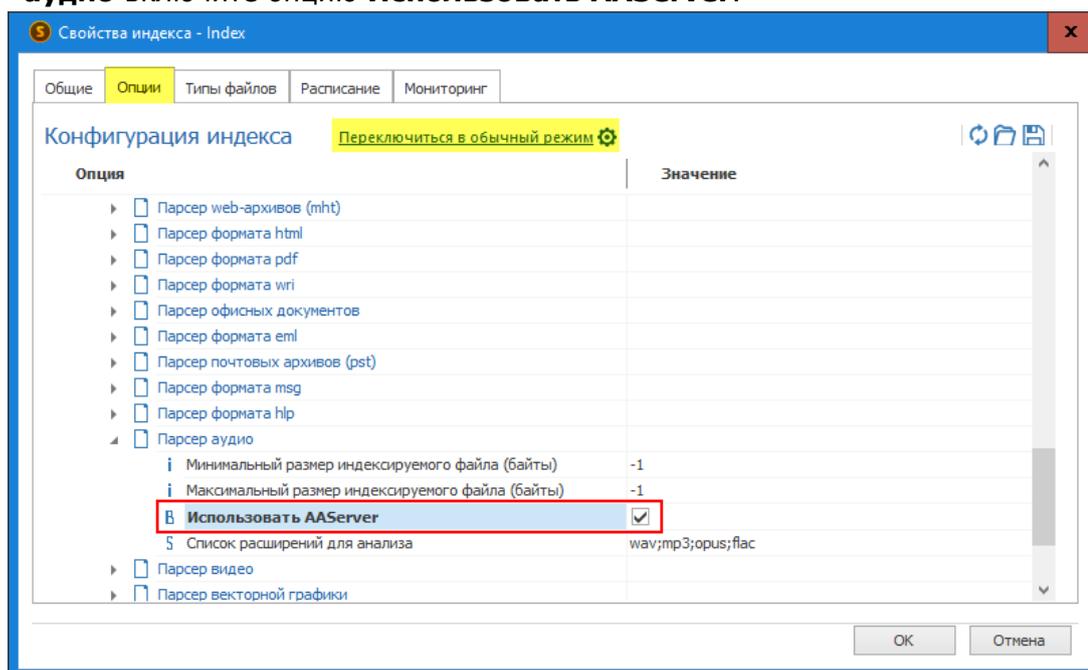


При необходимости добавьте расписание индексации для данного индекса и убедитесь, что установлен флажок **Запустить индексацию после создания индекса**.

Нажмите **Завершить**.

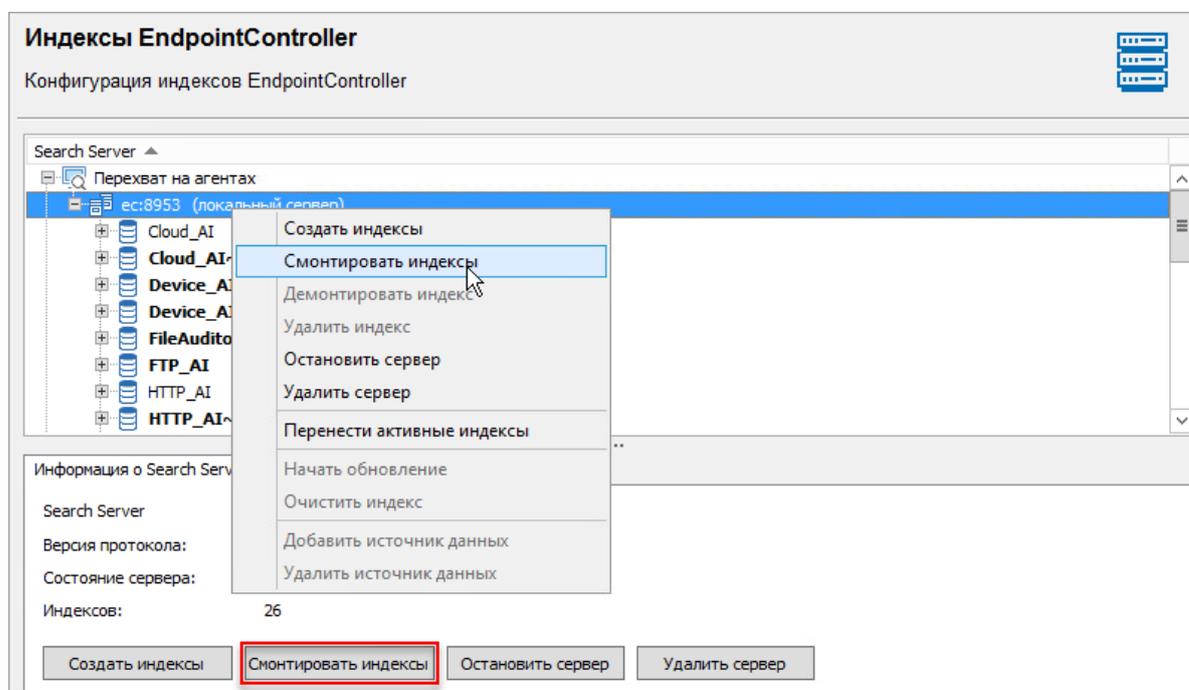


7. В свойствах выбранного/созданного индекса на вкладке **Опции** в группе **Парсер аудио** включите опцию **Использовать AAServer**.

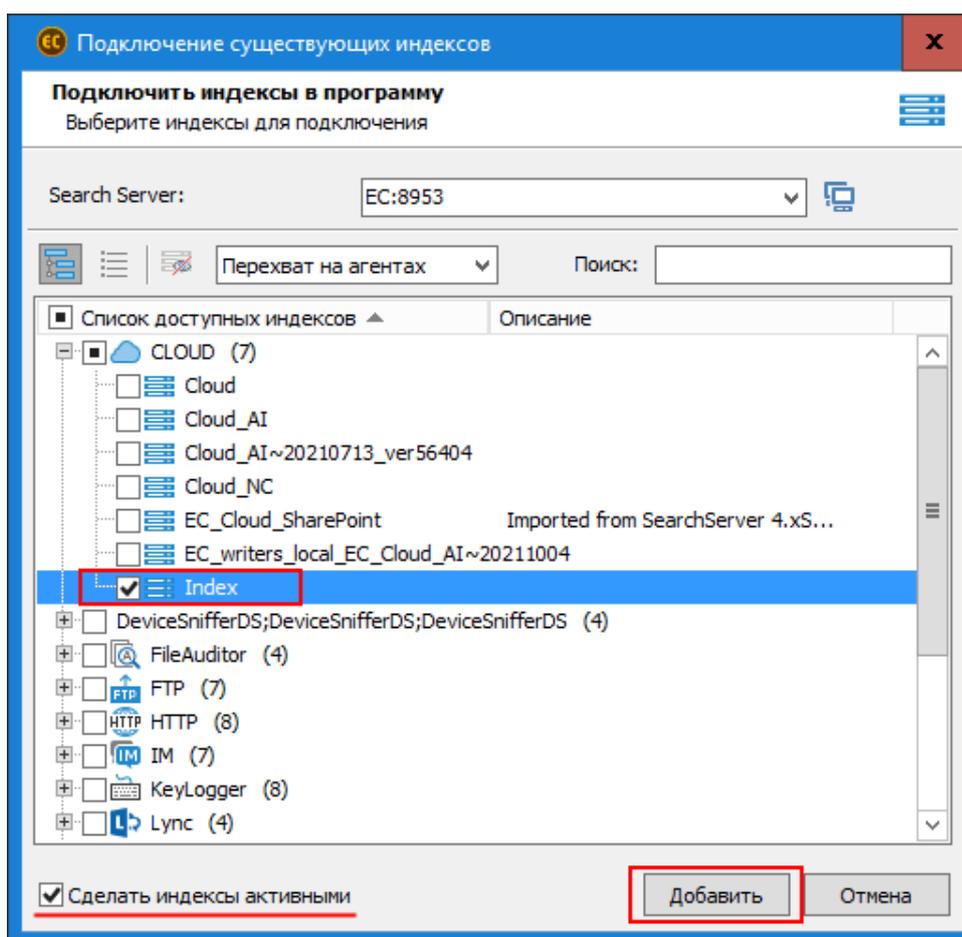


На этом необходимые настройки для распознавания речи с использованием движка Vosk в консоли SearchServer завершены.

8. После этого требуется подключить (смонтировать) созданный индекс в консоли EndpointController. Для этого перейдите в раздел SearchServer.
9. Выделите сервер, на котором установлен компонент SearchServer, и щелкните кнопку **Смонтировать индексы**. В качестве альтернативного варианта можно выбрать команду **Смонтировать индексы** из контекстного меню.



Отметьте флажком созданный ранее индекс. Чтобы перехват сохранялся в выбранном индексе, установите флажок **Сделать индексы активными**.



Смонтированные индексы будут отображены в окне просмотра.