

D DIGIT / РЕЧЕВОЙ ИНТЕЛЛЕКТ / 

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

«Digit Речевой интеллект»

Инструкция по установке и эксплуатации

Листов 8

СОДЕРЖАНИЕ

1	ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ	4
2	ТРЕБОВАНИЯ К СЕРВЕРНОМУ ОБОРУДОВАНИЮ	5
2.1	Требования к серверной части модуля «Digit Речевой интеллект».....	5
3	ИНСТРУКЦИЯ ПО УСТАНОВКЕ КОМПОНЕНТОВ ПРИЛОЖЕНИЯ В РУЧНОМ РЕЖИМЕ	6
3.1	Подготовка.....	6
3.2	Подготовка локальной среды	6
3.3	Установка Docker	6
3.4	Подготовка окружения	7
3.5	Развертывание сервисов ASR (ВМ 1)	7
3.6	Развертывание сервисов LLM (ВМ 2)	7
3.7	Настройка LLM-подключения	7
3.8	Проверка работоспособности.....	8

СПИСОК ТЕРМИНОВ И СОКРАЩЕНИЙ

Термин, сокращение	Значение
АРМ	Автоматизированное рабочее место – индивидуальный комплекс технических и программных средств, предназначенный для автоматизации профессионального труда специалиста и обеспечивающий подготовку, редактирование, поиск, отображение на экране и печать необходимых ему документов и данных
ПО	Программное обеспечение
ASR (Automatic Speech Recognition)	Технология автоматического распознавания речи, предназначенная для преобразования устной аудио- или видеозаписи в человеко-читаемый текст
VAD (Voice Activity Detection)	Технология цифровой обработки сигналов, предназначенная для обнаружения речи в непрерывном аудиопотоке и отделения её от фонового шума или тишины
LLM (Large Language Model)	Большая языковая модель, основанная на нейронной сети с множеством параметров (миллиарды весовых коэффициентов и более), которая проходит предварительное обучение на обширных массивах неразмеченного текста методами самообучения

1 ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Для развертывания программного обеспечения «Digit Речевой интеллект» программно-аппаратный комплекс должен иметь сервер приложений, функционирующие под управлением операционными системами Astra Linux SE.

Требования к операционным системам Astra Linux SE:

- Архитектура: x86_64.

2 ТРЕБОВАНИЯ К СЕРВЕРНОМУ ОБОРУДОВАНИЮ

2.1 Требования к серверной части модуля «Digit Речевой интеллект»

Рекомендуемые минимальные конфигурации серверного оборудования, а также перечень необходимого системного и прикладного программного обеспечения для серверной части модуля «Digit Речевой интеллект» приведены в таблице 1.

Таблица 1 - Требования к серверной части «Digit Речевой интеллект» (базовый функционал)

№ п/п	Наименование параметра	Значение параметра
Сервер приложений сервиса распознавания аудио и видеоматериалов, и сервиса обработки и распознавания речи		
1.	CPU - Количество ядер процессора, шт. (с поддержкой инструкций AVX, AVX2 или AVX512)	16
2.	RAM - Объем оперативной памяти, ГБ.	16
3.	HDD – Объем дисков, ГБ	40 (ОС) + 512 (ПО)
Операционная система:	«Astra Linux SE»	
ПО, которое будет установлено на сервер:	Redis, S3 хранилище MinIO, Docker Compose, Docker Engine	
Сервер приложений сервиса обработки и распознавания речи		
1.	CPU - Количество ядер процессора, шт. (с поддержкой инструкций AVX, AVX2 или AVX512)	16
2.	RAM - Объем оперативной памяти, ГБ.	16
3.	HDD – Объем дисков, ГБ.	40 (ОС) + 512 (ПО)
4.	GPU – Объем видеопамяти, ГБ. (рекомендуемый минимум)	40
Операционная система:	«Astra Linux SE»	
ПО, которое будет установлено на сервер:	Docker Compose, Docker Engine, vLLM	

3 ИНСТРУКЦИЯ ПО УСТАНОВКЕ КОМПОНЕНТОВ ПРИЛОЖЕНИЯ В РУЧНОМ РЕЖИМЕ

Для развертывания программного обеспечения «Digit Речевой интеллект» необходимо выполнить следующие действия в установленном порядке:

1. Подготовка сервера;
2. Установка Docker
3. Подготовка локальной среды;
4. Настройка окружения;
5. Развёртывание сервисов ASR;
6. Развёртывание сервисов LLM;
7. Проверка работоспособности системы.

3.1 Подготовка

1. Скачать архив из предоставленного источника.
2. Получить учетные данные доступа:
 - Username - *по запросу*
 - Password - *по запросу*
3. Скопировать архив:
`cp digit-voice.zip /tmp/`
4. Установить необходимые пакеты:
`sudo apt-get update`
`sudo apt-get install -y curl unzip`

3.2 Подготовка локальной среды

Разархивировать архив digit-voice.zip

```
cd /tmp/  
unzip digit-voice.zip  
cd digit-voice
```

3.3 Установка Docker

Все окружение будет развернуто в докере, поэтому сначала необходимо установить сам докер:

```
sudo apt-get update  
sudo apt-get install -y docker.io docker-compose-plugin  
sudo systemctl enable --now docker  
sudo groupadd docker || true  
sudo usermod -aG docker $USER
```

Проверка:

```
docker -v  
docker compose version
```

3.4 Подготовка окружения

Для настройки переменного окружения необходимо выполнить команды:

```
cd scai-spirit  
cp example.env .env
```

Настроить файл:

```
scai-speech-inference/config/config.yaml
```

Параметры:

```
version: "1.3.0"  
asr:  
  use_flash: false
```

3.5 Развертывание сервисов ASR (ВМ 1)

1. Загрузка образов

```
docker load -i redis.tar  
docker load -i minio.tar  
docker load -i scai-speech-inference.tar  
docker load -i scai-spirit.tar  
docker load -i scai-spirit-ui.tar
```

2. Запуск:

```
docker compose -p digit-voice -f docker-compose-asr.yaml up -d
```

3.6 Развертывание сервисов LLM (ВМ 2)

1. Загрузка образов:

```
docker load -i vllm.tar
```

2. Запуск:

```
docker compose -f docker-compose-llm.yaml up -d
```

3.7 Настройка LLM-подключения

1. Файл:

```
scai-spirit-ui/config/config.yaml
```

2. Настроить:

```
digit_voice:  
  log_level: DEBUG  
scai_spirit:  
  url: http://scai-spirit:8000  
  upload_endpoint: /api/v1/files/upload
```

```
recognize_endpoint: /api/v1/speech/async-recognize
max_retries: 70
interval: 10
llm:
timeout: 300.0
api_key: VLLM-PLACEHOLDER-API-KEY
model_name: /models/Qwen3-14B-FP8
base_url: http://<IP_VM_LLM>:8088
model_type: vllm
max_tokens: 8000
temperature: 0.1
```

3.8 Проверка работоспособности

1. Проверить, что все контейнеры запущены:

```
docker ps
```

Результат:

- все контейнеры запущены;
- статус **healthy**.

Таблица 2 - Соответствие сервисов и портов

№	Сервис	Порт	Назначение
1.	MinIO API	9000	объектное хранилище
2.	MinIO Console	9001	веб-интерфейс MinIO
3.	scai-spirit API	8000	backend API
4.	scai-speech-inference	8971	распознавание речи
5.	vLLM API	8088	LLM-модель
6.	Redis	6379	очередь задач