

Процессы по поддержанию жизненного цикла Сервиса инференса MarQus

Сервис инференса нейросетевых моделей MarQus для
автоматизации процессов распознавания и сортировки различных
типов ТБО.

2023

Содержание

Аннотация

1. Общие сведения
 - 1.1. Наименование
 - 1.2. Состав программных средств и используемый стек технологий
2. Поддержание жизненного цикла
 - 2.1. Назначение сопровождения сервиса
 - 2.2. Сервисные процессы сопровождения сервиса
3. Информация о персонале
 - 3.1. Уровень подготовки пользователей
 - 3.2. Персонал, обеспечивающий работу по интеграции сервиса с роботизированными линиями сортировки
 - 3.3. Персонал, обеспечивающий техническую поддержку и модернизацию
4. Дополнительная информация
 - 4.1. Информация о процессе разработки
 - 4.2. Информация о персонале, обеспечивающем поддержание жизненного цикла
 - 4.3. Информация о процессе сопровождения и совершенствования сервиса

Аннотация

Настоящий документ (далее – Описание) распространяется на программный продукт «сервиса инференса нейросетевых моделей для автоматизации процессов распознавания и сортировки различных типов ТБО Marqus» (далее – сервис).

Данное Описание содержит сведения о процессах, обеспечивающих поддержание жизненного цикла сервиса, а также информацию о персонале для устранения неисправностей, выявленных в ходе его эксплуатации.

В разделе «Общие сведения» указаны наименование сервиса, программное обеспечение, необходимое для функционирования сервиса, и перечень технологий, который используется в сервисе.

В разделе «Поддержание жизненного цикла системы» приведены сведения о процессах, обеспечивающих поддержание жизненного цикла сервиса.

В разделе «Информация о персонале» приведены сведения о персонале, обеспечивающем работу и модернизацию сервиса.

1. Общие сведения

1.1. Наименование

Наименование – Сервис инференса нейросетевых моделей для автоматизации процессов распознавания и сортировки различных типов ТБО MarQus.

1.2. Состав программных средств и используемый стек технологий

В основе архитектуры применяются популярные, открытые компоненты.

Библиотеки машинного обучения

- PyTorch – фреймворк машинного обучения для языка Python с открытым исходным кодом, созданный на базе Torch. Используется для решения различных задач: компьютерное зрение, обработка естественного языка.
- TorchVision - состоит из популярных наборов данных, архитектур моделей и общих преобразований изображений для компьютерного зрения.
- OpenCV – библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения с открытым кодом. Реализована на C/C++, также разрабатывается для Python, Java, Ruby, Matlab, Lua и других языков.
- Ultralytics - передовая современная модель (SOTA), основанная на успехе предыдущих версий YOLO и предлагающая новые функции и улучшения для дальнейшего повышения производительности и гибкости.

- DastAI - библиотека глубокого обучения, которая предоставляет специалистам-практикам высокоуровневые компоненты, которые могут быстро и легко обеспечивать самые современные результаты в стандартных областях глубокого обучения.
- LAP - библиотека решения задач линейного назначения, использующая алгоритм Йонкера-Волгенанта для плотных или разреженных матриц.

Общепотребительные библиотеки

- FastAPI – веб-фреймворк для создания API, написанный на Python. Один из самых быстрых и популярных веб-фреймворков, написанных на Python.
- Hypercorn – это веб-сервер ASGI. Hypercorn поддерживает спецификации HTTP/1, HTTP/2, WebSockets (через HTTP/1 и HTTP/2), ASGI/2 и ASGI/3. Hypercorn может использовать рабочие типы asyncio, uvloop или trio.
- Dynaconf – управление настройками в проекте.
- NumPy – библиотека с открытым исходным кодом для языка программирования Python. Поддержка многомерных массивов, поддержка высокоуровневых математических функций, предназначенных для работы с многомерными массивами.
- Pandas – библиотека на языке Python для обработки и анализа данных.
- imageZMQ – набор классов Python для передачи изображения OpenCV с помощью обмена сообщениями PyZMQ.
- Protocol Buffers – протокол сериализации структурированных данных, бинарная альтернатива текстовому формату XML.

При разработке сервиса использован язык программирования Python.

Для обеспечения унифицированного предоставления данных разработан слой REST API. Для передачи приёма и передачи изображений используется ZeroMQ - высокопроизводительная асинхронная библиотека обмена сообщениями, ориентированная на использование в распределённых и параллельных вычислениях.

Для развёртывания системы необходимы следующие предустановки:

- Операционная система: Любой дистрибутив Linux с ядром 5.4 или выше
- CPU: Intel(R) Core(TM) i7-7700K CPU @ 4.20GHz and newer
- RAM: минимум 8 GB
- HDD/SSD: минимум 64 GB
- GPU: Nvidia 1080Ti или новей
- GPU: память не ниже 11264 МБ
- Compiler: GCC 7.3.1 или новей
- CUDA: 11.0 или новей
- Docker: 20.10 или новей

2. Поддержание жизненного цикла сервиса

Поддержание жизненного цикла сервиса осуществляется за счет сопровождения.

2.1. Назначение сопровождения сервиса

Сопровождение сервиса позволяет: - обеспечить отсутствие простоя в работе пользователей по причине невозможности функционирования сервиса (аварийная ситуация, ошибки в работе сервиса, ошибки при развёртывании и т.п.); - обеспечить гарантию корректного функционирования сервиса; - обеспечить гарантию дальнейшего развития функционала сервиса.

2.2. Сервисные процессы сопровождения

Для обеспечения жизненного цикла при наличии договора поддержки в сопровождение сервиса включены следующие сервисные процессы: - консультирование пользователей/администраторов сервиса по вопросам эксплуатации по телефону, электронной почте или письменно по запросу Заказчика; - обеспечение Заказчика новыми версиями сервиса по мере их выхода; - обеспечение Заказчика изменениями и дополнениями к эксплуатационной документации; - устранение ошибок в случае их выявления при работе с сервисом.

Техническая поддержка пользователей.

Техническая поддержка пользователей осуществляется в формате консультирования пользователей/администраторов сервиса по вопросам установки, переустановки, администрирования и эксплуатации программного обеспечения по электронным каналам связи (телефону, электронной почте) или письменно по запросу.

В рамках технической поддержки сервиса оказываются следующие услуги:

- помощь в установке программного обеспечения;
- помощь в настройке и администрировании программного обеспечения;
- помощь в установке обновлений программного обеспечения;
- помощь в поиске и устранении проблем в случае некорректной установки обновления программного обеспечения;
- пояснение функциональности программного обеспечения, помощь в эксплуатации;
- предоставление актуальной документации по установке, настройке и работе программного обеспечения;
- общие консультации по выбору аппаратного обеспечения для высокой производительности работы сервиса.

В рамках модернизации оказываются следующие услуги:

- выявление ошибок в функционировании сервиса;
- исправление ошибок, выявленных в функционировании сервиса;

- прием заявок от Заказчика на внесение изменений и дополнений в сервис;
- согласование с Заказчиком возможности и сроков исполнения заявок, оказание консультативной помощи по вопросам технической реализации пожеланий, указанных в заявке;
- модернизация сервиса по заявкам Заказчика;
- модернизация сервиса в связи с изменением ее функциональных составляющих;
- предоставление Заказчику новых версий сервиса, выпущенных в результате модернизации и исправления ошибок;
- обеспечение пользователей изменениями и дополнениями к эксплуатационной документации;
- предоставление Заказчику неисключительных прав на использование новых версий сервиса, выпущенных в результате модернизации и исправления ошибок.

3. Информация о персонале

3.1. Уровень подготовки пользователей сервиса

Пользователи сервиса должны иметь опыт работы с:

- Docker;
- ОС семейства Linux;
- любым языком программирования встраиваемых систем. Предпочтительно C/C++, Java или Python.

Пользователь должен обладать следующими навыками:

- владеть навыками разработчика на уровне интеграций готовых модулей;
- знать бизнес-процессы, связанные с профильной деятельностью.

Перед началом использования сервиса пользователь обязан пройти обучение, ознакомиться с должностными инструкциями и пользовательской документацией.

3.2. Персонал, обеспечивающий работу по интеграции сервиса с роботизированными линиями сортировки

Разработчики, использующие сервис, должны обладать навыками работы с персональным компьютером на уровне разработчика встраиваемых систем. Для работы с сервисом разработчику необходимо изучить свои должностные инструкции и руководства интегратора сервиса, а также пройти курс обучения по сервису. Обязательно знание работы вычислительной техники и программного обеспечения в локальных сетях, опыт работы с ОС семейства Linux, навыков программирования.

3.3. Персонал, обеспечивающий техническую поддержку и модернизацию

Специалисты, обеспечивающие техническую поддержку и развитие сервиса, должны обладать следующими знаниями и навыками:

- владение персональным компьютером на уровне продвинутого разработчика;

- знание функциональных возможностей сервиса и особенностей работы с ними;
- знания языков Python, C/C++;
- знания протоколов TCP, HTTP, умение пользоваться инструментами разработчика и удалённой отладки;
- владение консолью Linux (unix-shell) на продвинутом уровне;
- Знание технологий контейнеризации, оркестрации контейнеров (Docker, Kubernetes);
- Глубоко понимать основы машинного обучения.

Специалисты, осуществляющие модернизацию сервиса, помимо вышеперечисленного, должны иметь квалификацию “аналитик данных”.

4. Дополнительная информация

4.1. Информация о процессе разработки

Данные о персонале, задействованном в процессе разработки:

- аналитик данных – 2 человека;
- инженер по машинному обучению – 1 человек;
- ведущий бизнес аналитик – 1 человек;
- backend разработчик – 1 человек;
- frontend разработчик – 1 человек;
- тестировщик – 1 человек.

Указанные специалисты являются штатными сотрудниками ООО «Зебрейнс».

Фактический адрес размещения инфраструктуры разработки:

- г.Краснодар, ул. Красная 69 (сервер с 2 x GPU GTX 4090)
- г.Ульяновск, ул. Федерации 86 (сервер с 1 x GPU GTX 3060)
- г.Москва, Selectel (сервер с 2 x GPU A5000)

Фактический адрес разработчиков:

- г.Краснодар, ул. Красная 69
- г.Ульяновск, ул. Федерации 86

4.2. Информация о персонале, обеспечивающем поддержание жизненного цикла

Данные о персонале, обеспечивающем поддержание жизненного цикла:

- аналитик данных – 2 человека;
- инженер по машинному обучению – 1 человек;
- ведущий бизнес аналитик – 1 человек;
- backend разработчик – 1 человек;
- frontend разработчик – 1 человек;
- тестировщик – 1 человек;
- специалист технической поддержки – 2 человека;

Указанные специалисты являются штатными сотрудниками ООО «Зебрейнс».

4.3. Информация о процессе сопровождения и совершенствования программного обеспечения сервиса

Средства коммуникации со службой сервисной поддержки:

- почта: marqus@zebrains.team
- режим работы сервисной поддержки: 8 часов в сутки 5 дней в неделю;
- режим работы технической поддержки: 8 часов в сутки 5 дней в неделю;
- количество персонала, задействованного в процессе сопровождения: 2 человека.

Средства совершенствования программного обеспечения:

- эволюционирование кодовой базы в рамках систематического преобразования и совершенствования, в соответствии с потребностями пользователей в рамках продукта;
- формирование пользовательских историй;
- получение обратной связи от пользователей;
- тестирование и выпуск релизов;
- поставка периодических обновлений и релизов программного обеспечения в рамках сервисной поддержки;
- сбор статистики, формирование датасетов и дообучение нейросетевых моделей.