

ИНТЕПРЕТИРУЕМЫЙ "MACHINE INTELLIGENCE"

---

D2PL = Distributed Dataflows PL

Язык программирования D2PL и компилятор для кластеризуемой массивно-параллельной виртуальной машины

---

Прозоров Александр, Лаборатория машинного интеллекта МФТИ

14 ноября 2019

ООО «РТЛаб»

[info@rtlab.ru](mailto:info@rtlab.ru)

[www.cybersyn.ch](http://www.cybersyn.ch)

## Предпосылки\* (1 из 3)

1

“Как сделать программу, которая  
пишет другие программы, которые  
пишут другие программы, которые  
пишут...”

# Предпосылки\* (1 из 3)



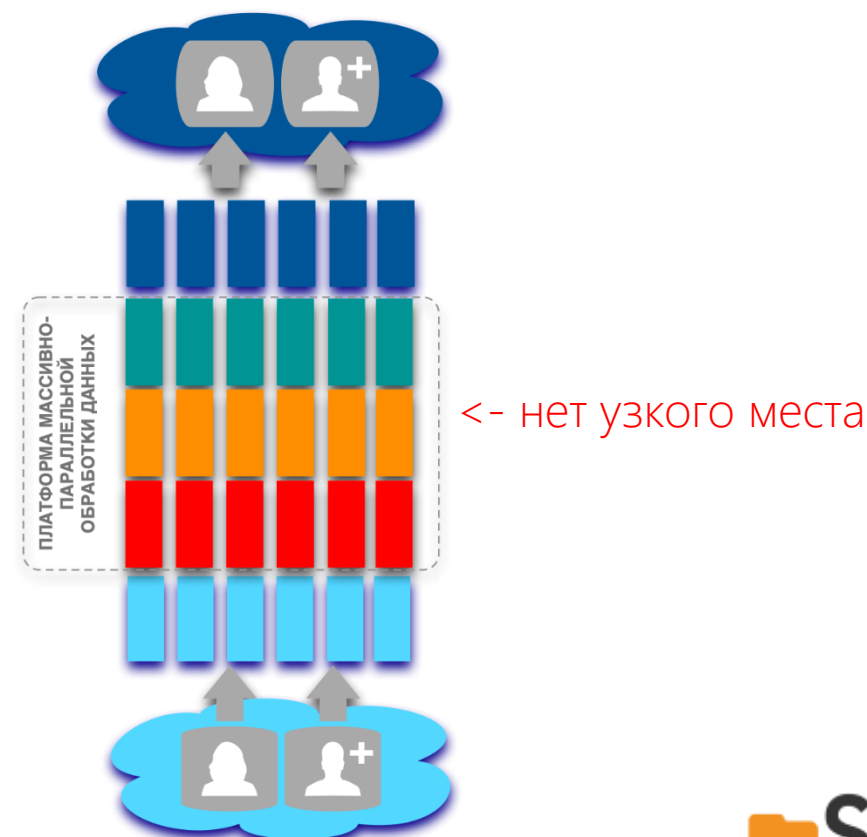
\*Источник: Лекция А. Прозорова для студентов МФТИ, "BigData на охране здоровья", 2014

# Предпосылки\* (2 из 3)

## «Традиционные» ИТ-системы



## Динамические МП ИТ-системы



\*Источник: Лекция А. Прозорова для студентов МФТИ, "BigData на охране здоровья", 2014

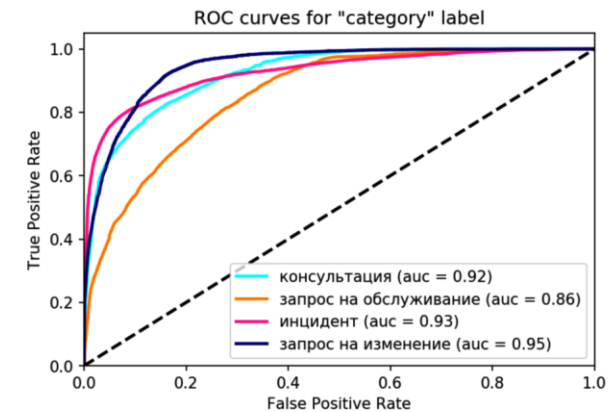
# Фундаментальная проблема ML-centric подхода

4

- ★ ML-centric подход – это построение систем машинного интеллекта на основе больших данных и стандартных вычислительных/ статистических методов
- ★ Корректная интерпретация результатов этих систем возможна только в рамках
  - ★ математических («вероятностных») метрик
  - ★ тестового набора данных

## Оценка качества бинарной классификации

- Классификатор выдаёт **score**
- Объекты упорядочиваются по **score**
- Выбор порога принятия решения
- **TPR, FPR** – true, false positive rate
- **AUC** – площадь под кривой, **доля правильно упорядоченных объектов**



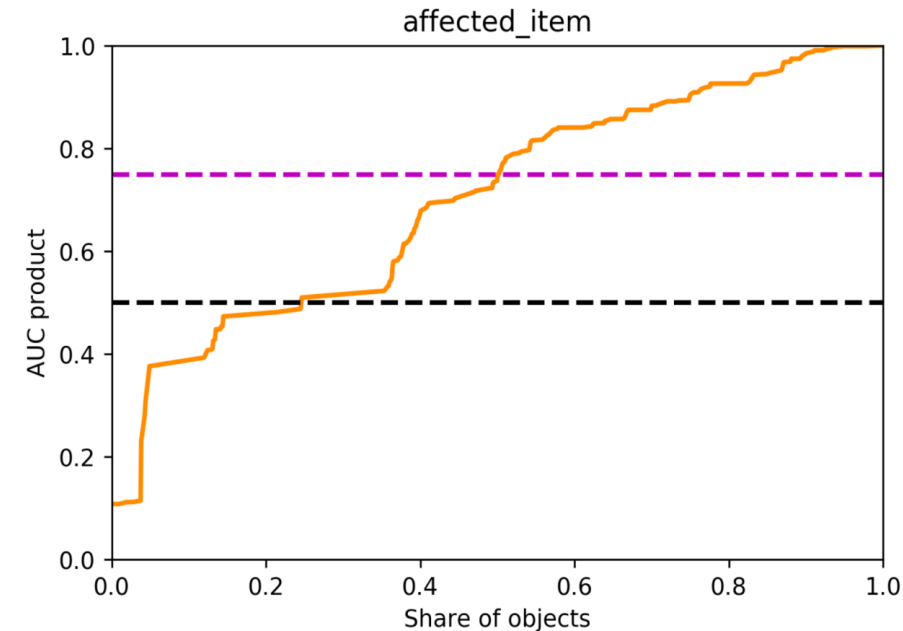
Оценка качества данных по различным типам обращений пользователей

**Вероятностный результат не гарантирует однозначности**

**В то время как заказчик формулирует задачу в виде системы логических правил и ожидает однозначности результата**

## ML-centric подход: плюсы

- ★ Доказанная результативность (поиск, игры, виртуальные помощники, etc)
- ★ Существуют развитые курсы обучения
- ★ Хорошо известен в профессиональном сообществе
- ★ Проработана инфраструктура разработки, эксплуатации и DevOps

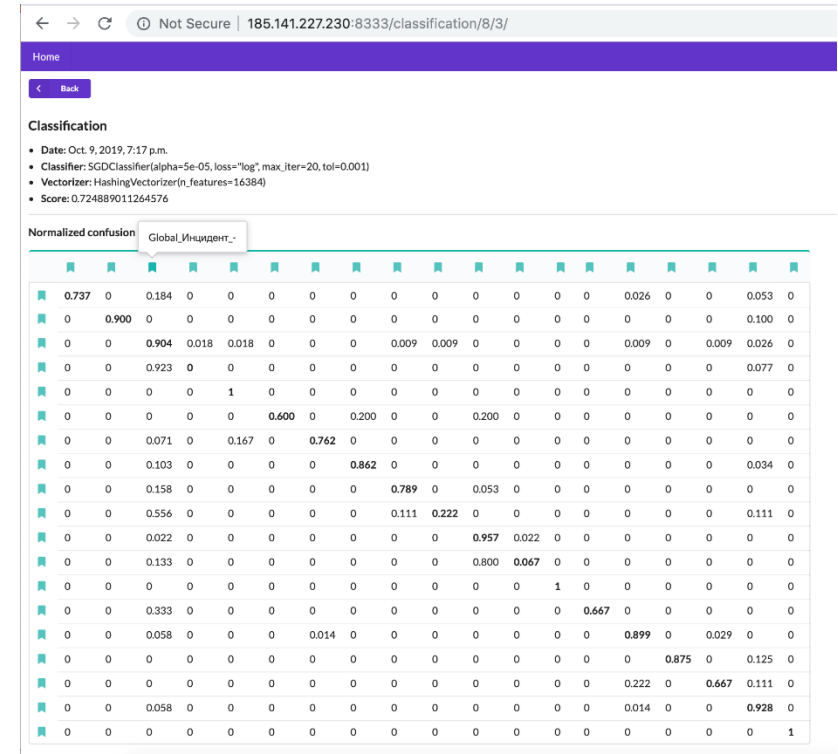


Пример анализа качества запросов по параметру affected\_item

# ML-centric подход: минусы

6

- ★ Далеко не везде и всегда есть BigData!
- ★ Много «green-field» отраслей и проектов
- ★ Сложно переносить в численные модели компетенции из регламентов и спецификаций бизнес-процессов
- ★ Проект занимает слишком много времени и ресурсов
  - ★ дата-инженерам нужно погружаться в предметную область
  - ★ балансировка дата-сетов делает этап подготовки/обучения длительным
  - ★ нужны глубокие теоретические знания и опыт в области статистики и вычислительных методов
  - ★ сложность архитектуры системы требует специалистов в различных областях (необходима большая команда)



Classification

- Date: Oct. 9, 2019, 7:17 p.m.
- Classifier: SGDClassifier(alpha=5e-05, loss='log', max\_iter=20, tol=0.001)
- Vectorizer: HashingVectorizer(n\_features=16384)
- Score: 0.724889011264576

Normalized confusion

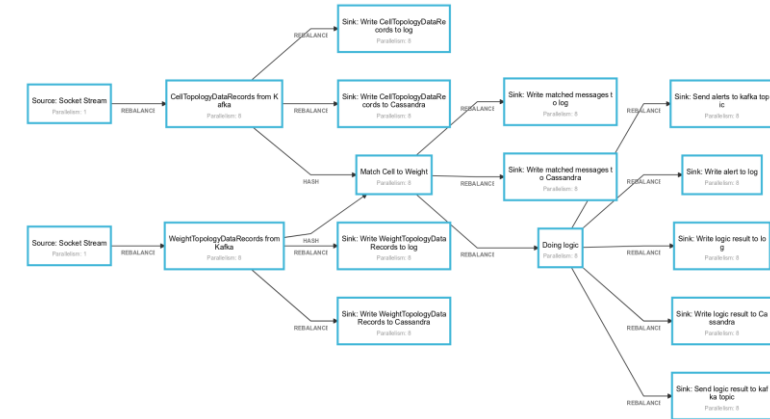
|       |       |       |       |       |       |       |       |       |       |       |   |   |   |   |       |       |       |       |       |   |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---|---|---|---|-------|-------|-------|-------|-------|---|
| 0.737 | 0     | 0.184 | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0.026 | 0     | 0     | 0.053 | 0 |
| 0     | 0.900 | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0.100 | 0 |
| 0     | 0     | 0.904 | 0.018 | 0.018 | 0     | 0     | 0     | 0.009 | 0.009 | 0     | 0 | 0 | 0 | 0 | 0.009 | 0     | 0.009 | 0.026 | 0     | 0 |
| 0     | 0     | 0.923 | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0.077 | 0 |
| 0     | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0     | 0 |
| 0     | 0     | 0     | 0     | 0     | 0.600 | 0     | 0.200 | 0     | 0     | 0.200 | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0     | 0 |
| 0     | 0     | 0.071 | 0     | 0.167 | 0     | 0.762 | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0     | 0 |
| 0     | 0     | 0.103 | 0     | 0     | 0     | 0     | 0.862 | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0.034 | 0 |
| 0     | 0     | 0.158 | 0     | 0     | 0     | 0     | 0     | 0.789 | 0     | 0.053 | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0     | 0 |
| 0     | 0     | 0.556 | 0     | 0     | 0     | 0     | 0     | 0.111 | 0.222 | 0     | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0.111 | 0 |
| 0     | 0     | 0.022 | 0     | 0     | 0     | 0     | 0     | 0     | 0.957 | 0.022 | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0     | 0 |
| 0     | 0     | 0.133 | 0     | 0     | 0     | 0     | 0     | 0     | 0.800 | 0.067 | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0     | 0 |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 1     | 0     | 0     | 0     | 0     | 0 |
| 0     | 0     | 0.333 | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0.667 | 0     | 0     | 0     | 0     | 0 |
| 0     | 0     | 0.058 | 0     | 0     | 0     | 0.014 | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0.899 | 0     | 0.029 | 0     | 0 |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0     | 0.875 | 0     | 0.125 | 0 |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0.222 | 0     | 0.667 | 0.111 | 0 |
| 0     | 0     | 0.058 | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0.014 | 0     | 0     | 0.928 | 0 |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0 | 0 | 0 | 0 | 0     | 0     | 0     | 0     | 0     | 1 |

Confusion matrix - оценка качества работы ML-классификатора

# Какое решение мы предлагаем?

7

- ★ «Интерпретируемый "Machine Intelligence"»
- ★ Для этого объединяем сильные стороны классического Computer Science и ML-centric подхода
- ★ Чтобы обеспечить интерпретацию результатов работы системы в целом, ML-задачи решаются на нижнем логическом уровне (уровень виртуальной машины)
- ★ При этом верхний логический уровень является согласованной D2PL-программой, содержащей декларативное описание алгоритма решения задач (уровень управления)
- ★ Уровень управления (D2PL-программы) гарантированно интерпретируется в форме блок-схем



```

description:
repo: "https://gitlab.rtlab.ru/va-platform/flink.git"
class: "rtlab.va.flink.shop.job.TopologyRecordsRouter"
purpose: Триггер производит роутинг входящих сообщений в правильный websocket
inputs:
- Получает сообщения из одного или более топиков
outputs:
- Отправляет принятые сообщения в websocket
alerts:
- Отправляет события ALT_SOCKET_ERROR при какой-либо ошибке сокета
environment:
- Redis
- Kafka
- td-agent
conditions:
- Работа триггера предполагает предварительную загрузку данных о

```

Граф исполнения и пример кода на d2pl

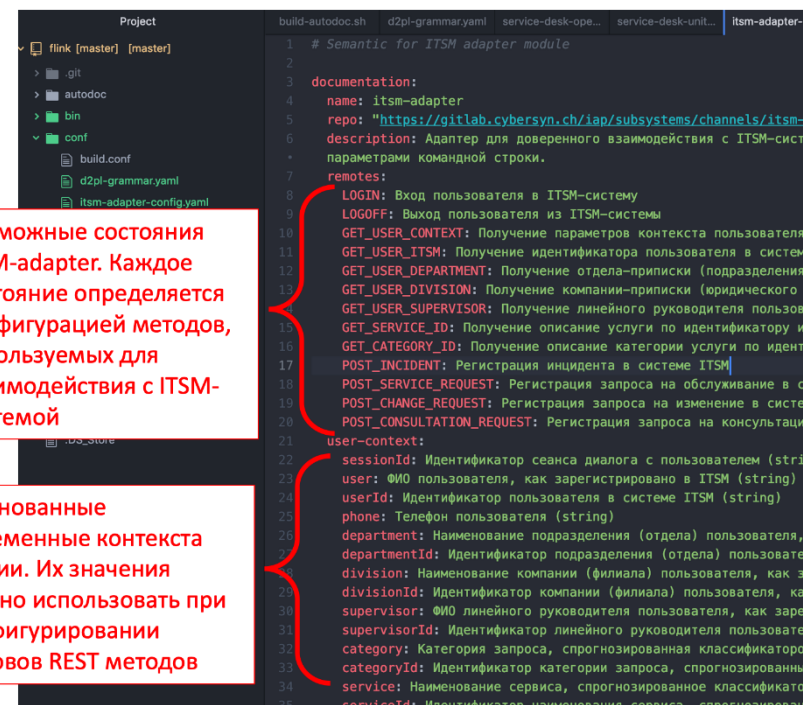


# Как это возможно?

8

- ★ Предлагаем новый подход к решению гиперкомплексных задач автоматического управления, подкрепленный новым типом вычислительных архитектур
- ★ Для этого разрабатываем
  - ◆ методологию решения задач автоматического управления
  - ◆ тип вычислительных архитектур (динамические массивно-параллельные системы)
- ★ «Интерпретируемый "Machine Intelligence"» в одинаковой степени опирается на обе эти составляющие

## Интеграция авторегистратора с ITSM



```

1 # Semantic for ITSM adapter module
2
3 documentation:
4   name: itsm-adapter
5   repo: "https://gitlab.cybersyn.ch/iap/subsystems/channels/itsm-adapter"
6   description: Адаптер для доверенного взаимодействия с ITSM-системой по
7     параметрам командной строки.
8   remotes:
9     LOGIN: Вход пользователя в ITSM-систему
10    LOGOFF: Выход пользователя из ITSM-системы
11    GET_USER_CONTEXT: Получение параметров контекста пользователя
12    GET_USER_ITSM: Получение идентификатора пользователя в системе
13    GET_USER_DEPARTMENT: Получение отдела-приписки (подразделения)
14    GET_USER_DIVISION: Получение компании-приписки (юридического лица)
15    GET_USER_SUPERVISOR: Получение линейного руководителя пользователя
16    GET_SERVICE_ID: Получение описание услуги по идентификатору или
17    GET_CATEGORY_ID: Получение описание категории услуги по идентификатору
18    POST_INCIDENT: Регистрация инцидента в системе ITSM
19    POST_SERVICE_REQUEST: Регистрация запроса на обслуживание в системе
20    POST_CHANGE_REQUEST: Регистрация запроса на изменение в системе
21    POST_CONSULTATION_REQUEST: Регистрация запроса на консультацию
22  user-context:
23    sessionId: Идентификатор сеанса диалога с пользователем (string)
24    user: ФИО пользователя, как зарегистрировано в ITSM (string)
25    userId: Идентификатор пользователя в системе ITSM (string)
26    phone: Телефон пользователя (string)
27    department: Наименование подразделения (отдела) пользователя,
28    departmentId: Идентификатор подразделения (отдела) пользователя,
29    division: Наименование компании (филиала) пользователя, как за
30    divisionId: Идентификатор компании (филиала) пользователя, как за
31    supervisor: ФИО линейного руководителя пользователя, как зарег
32    supervisorId: Идентификатор линейного руководителя пользователя
33    category: Категория запроса, спрогнозированная классификатором
34    categoryId: Идентификатор категории запроса, спрогнозированный
35    service: Наименование сервиса, спрогнозированное классификатором
36    serviceId: Идентификатор наименования сервиса, спрогнозированный
  
```

**Возможные состояния ITSM-adapter. Каждое состояние определяется конфигурацией методов, используемых для взаимодействия с ITSM-системой**

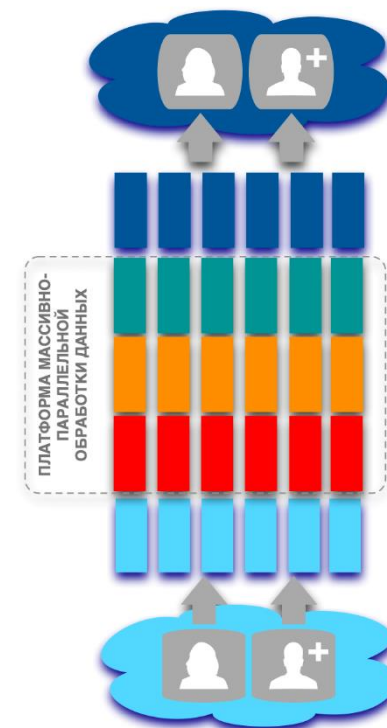
**Именованные переменные контекста сессии. Их значения можно использовать при конфигурировании вызовов REST методов**

# Структура нового поколения вычислительных архитектур

9

- ★ Язык программирования D2PL (машинное представление иерархии бизнес-процессов)
- ★ Распределенная по сети виртуальная машина D2PL (собираем из готовых компонентов и используем как есть)
- ★ Среда разработки D2PL в составе: IDE, компилятор для VM, отладчик уровня исходного кода (разрабатываем на основе готовых компонент)
- ★ Среда непрерывной интеграции и управления релизами, мониторинга их безопасности и технической поддержки (DevSecOps), собираем из готовых компонентов и интегрируем со средой разработки
- ★ Транслятор из формальной нотации описания бизнес-процессов типа IDEF\*/ARIS в программы на D2PL, разрабатываем и интегрируем со средой разработки

## Динамические МП ИТ-системы

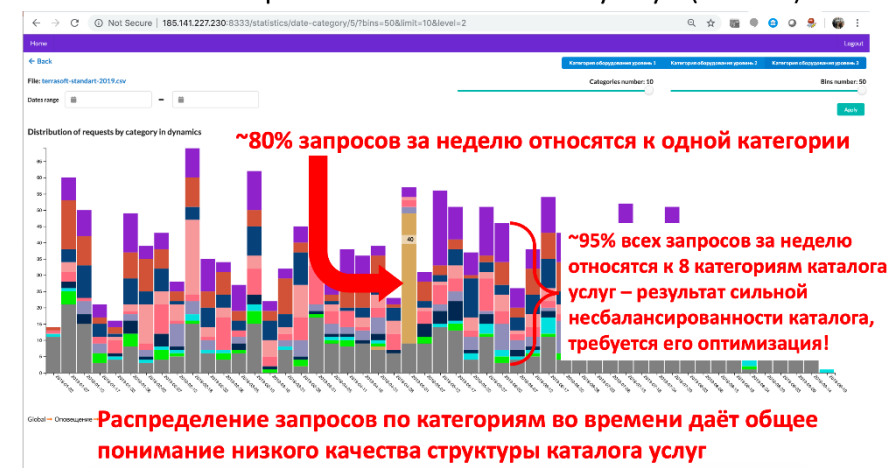


# Особенности новой методологии решения задач АУ

10

- ★ Вводим новые понятия и уточняем старые
- ★ Используем сеть «семантических ячеек» для описания структуры (онтологии) предметной области
- ★ Используем сеть «бизнес-процессов» для описания динамики (жизненного цикла) предметной области
- ★ Используем агентно-ориентированный подход для формализации решений задач АУ
- ★ Применяем сквозной жизненный цикл релизов (проектирование, разработка, развертывание, эксплуатация, техническая поддержка и вывод из эксплуатации)

## Выявление проблем с каталогом услуг (3 из 5)

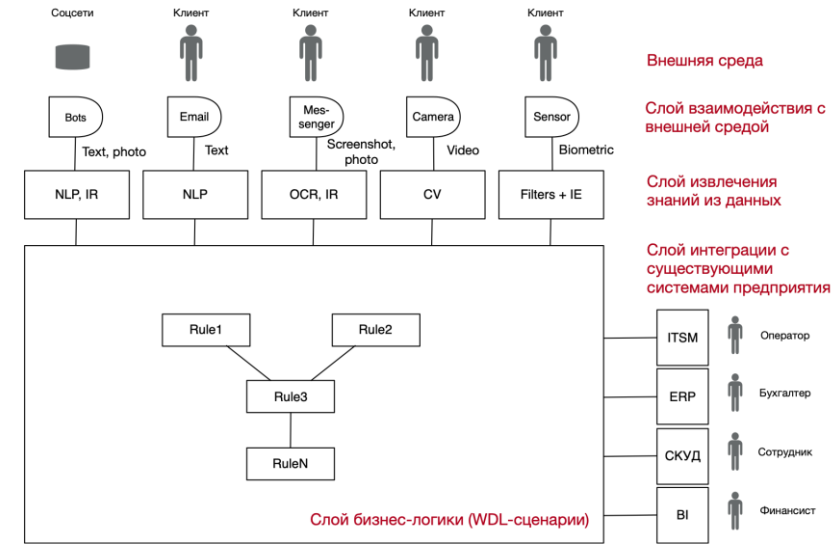


Внешний вид интерфейса пользователя контура подготовки моделей. На примере – анализ выгрузки из ITSM-системы хорошо известной компании

# Свойства новой методологии решения задач АУ

11

- ★ Интерпретация результатов работы системы в терминах предметной области
- ★ Работает там, где есть данные, но нет BigData (то есть информация о предметной области, содержащаяся в накопленном массиве данных, «не сбалансирована»)
- ★ Работает в «green-field» отраслях и проектах (то есть там, где надо запускаться с «0» - совсем нет или не могут быть накоплены данные с информацией о предметной области)
- ★ Требуется значительно меньшее время, необходимое для реализации проектов, чем в ML-centric подходе

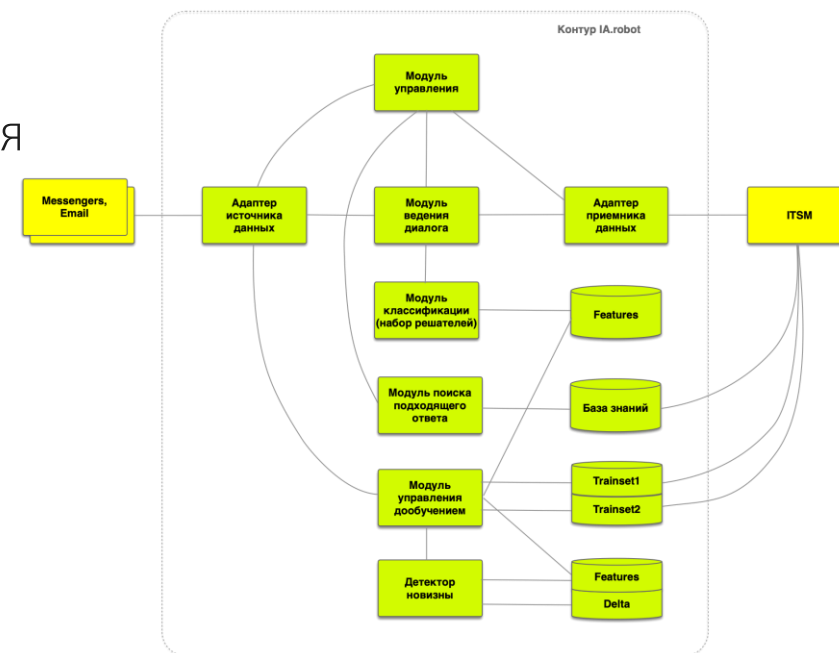


Версия функциональной модели платформы от 2015 года

# Почему требуется малое время для решения задач АУ?

12

- ★ Используются существующие описания предметной области в виде нотаций типа IDEF\*/ARIS
- ★ Экспертов в предметной области привлекают для описания бизнес-процессов в формальной нотации IDEF\*/ARIS на основе существующих регламентов
- ★ Описания бизнес-процессов в формальной нотации IDEF\*/ARIS транслируются в программы на языке D2PL
- ★ ML-задачи решаются в виде стандартных компонентов виртуальной машины (Encoder, Processor, Automator, etc)
- ★ Происходит пополнение библиотеки стандартных компонентов, как следствие, существенно увеличивается эффект от наиболее сильных людей в области вычислительных/статистических методов и системного программирования

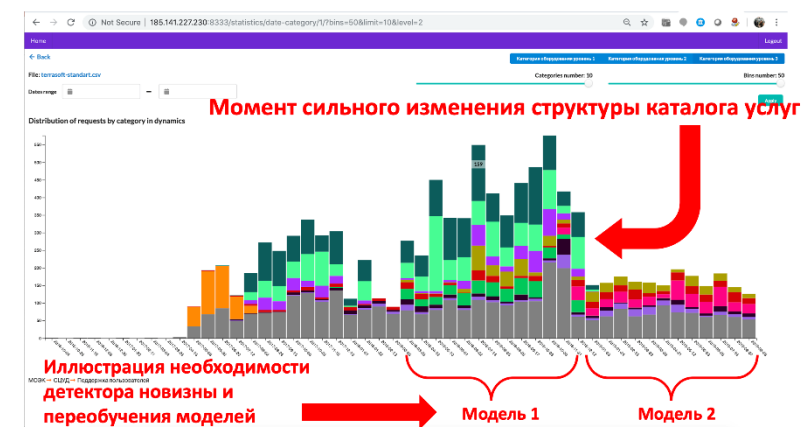


Функциональная модель виртуального помощника для задач ITSM

## Какие задачи АУ уже отработаны?

- ★ Видео-СКУД (автоматическое открытие/закрытие дверей + автоматический ввод актуальных фото в систему + интеграция с существующим СКУД)
- ★ Магазин-робот (продажи + логистика)
- ★ Самоактуализируемая база знаний отелей и достопримечательностей (world wide)
- ★ Виртуальный агент по регистрации инцидентов и заявок, получение справок типа 2НДФЛ, статей по базе знаний, автоответы по базе конфигурационных единиц (ITSM + back office)
- ★ А/Б тестирование упаковки/раскладки в товаров магазине (на основе замеров статистики направления взгляда респондентов)
- ★ Визуальная дефектоскопия штампованных стальных деталей кузовов для машин Renault/Nissan
- ★ Целеуказание крановщику и отслеживание фактов достижения целей
- ★ Масштабный сбор данных с бесконтактных датчиков с целью постоянного мониторинга состояния здоровья (проект eNurse)

### Выявление проблем с каталогом услуг (5 из 5)

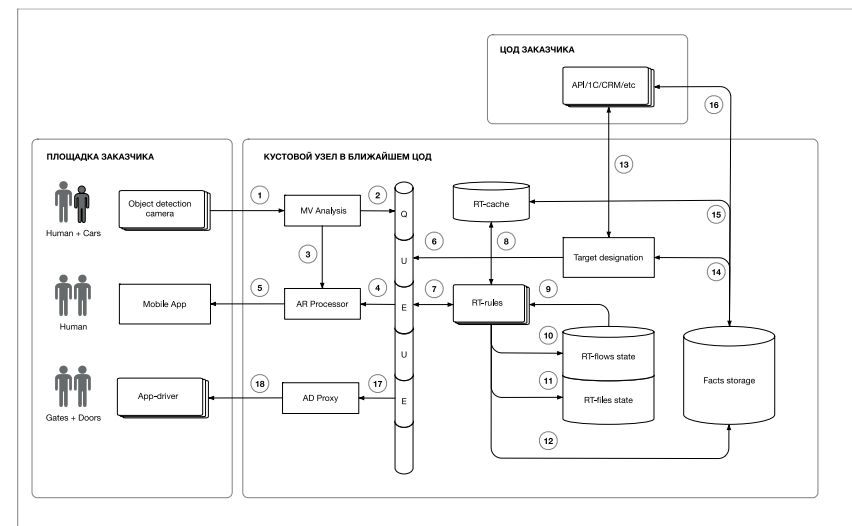


Внешний вид интерфейса пользователя контура подготовки моделей. На примере – анализ выгрузки из ITSM-системы хорошо известной компании

## Куда движемся?

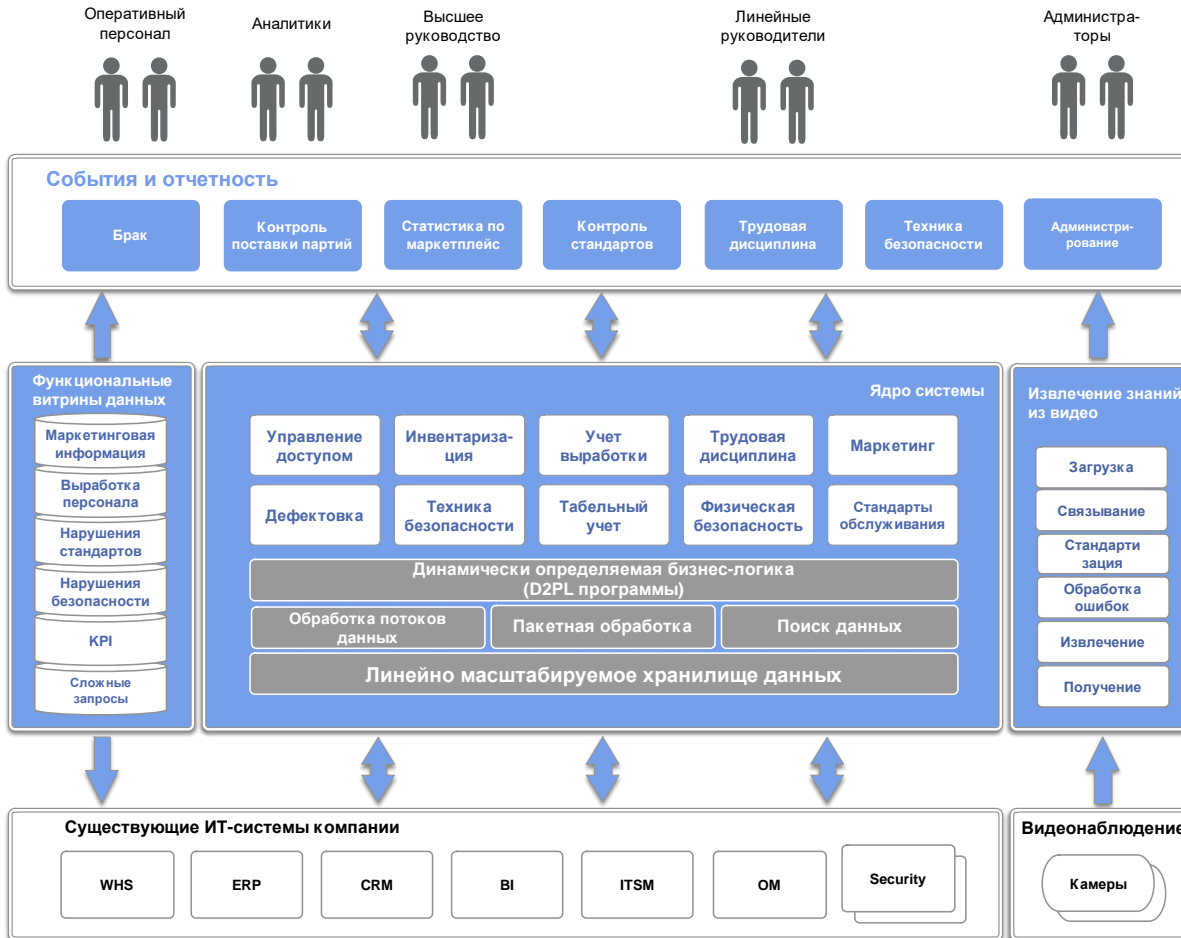
14

- ★ Развиваем прикладные конфигурации платформы (виртуальный агент, дефектоскопия, целеуказание, база знаний)
- ★ Развиваем DevSecOps среду (замкнутый жизненный цикл релизов, включая техническую поддержку и ТОиР)
- ★ Поэтапно реализуем компилятор D2PL
- ★ Формируем требования к IDE и отладчику уровня исходного кода
- ★ Ищем единомышленников и волонтеров («посыл во вселенную»)



Пример конфигурации платформы для задачи единой СКУД для жителей коттеджного поселка

# Приложение 1: Функциональная архитектура\*



Пользователи

Слой представления данных для пользователей

Массивно-параллельные вычисления

ИТ-ландшафт заказчика



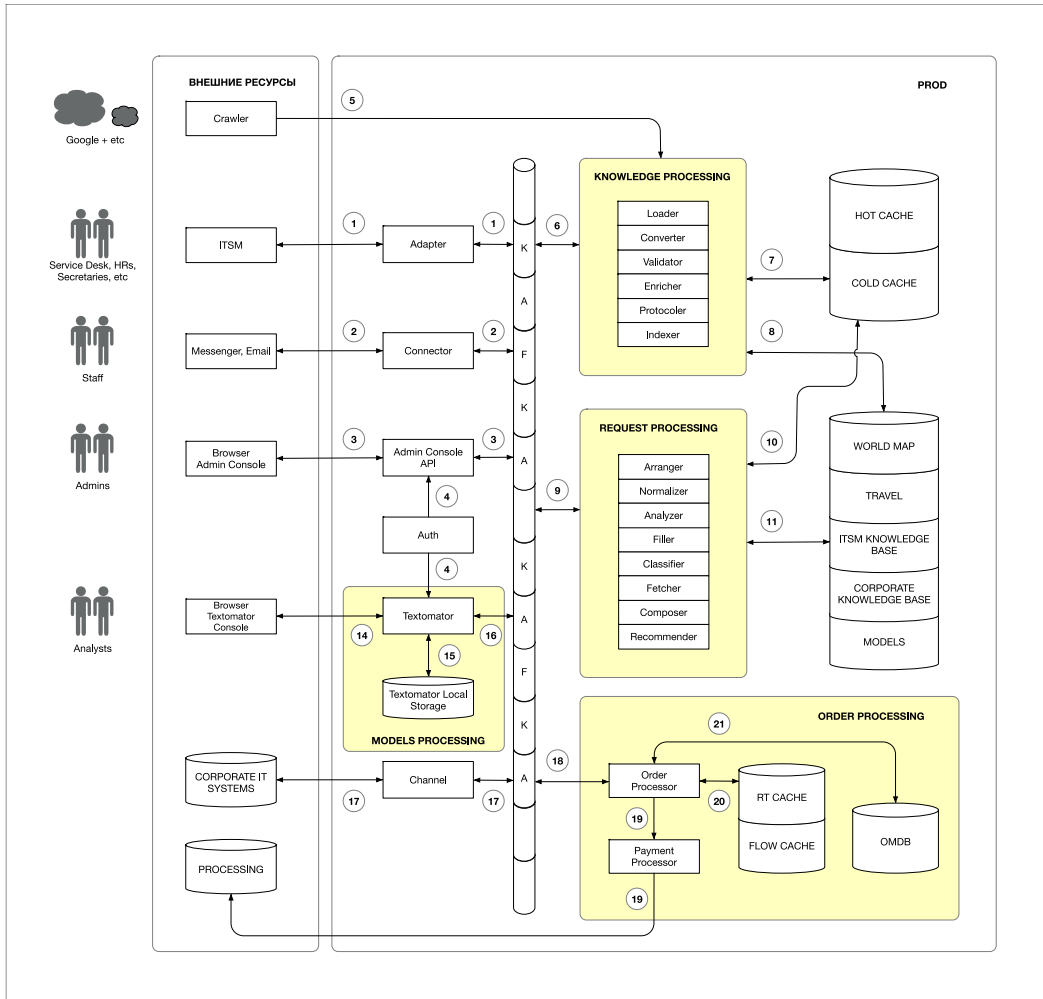
Сотрудники



\*На примере проекта: конфигурация платформы в режиме массивной обработки видео с камер наблюдения



# Приложение 2: Архитектура виртуальной машины\*

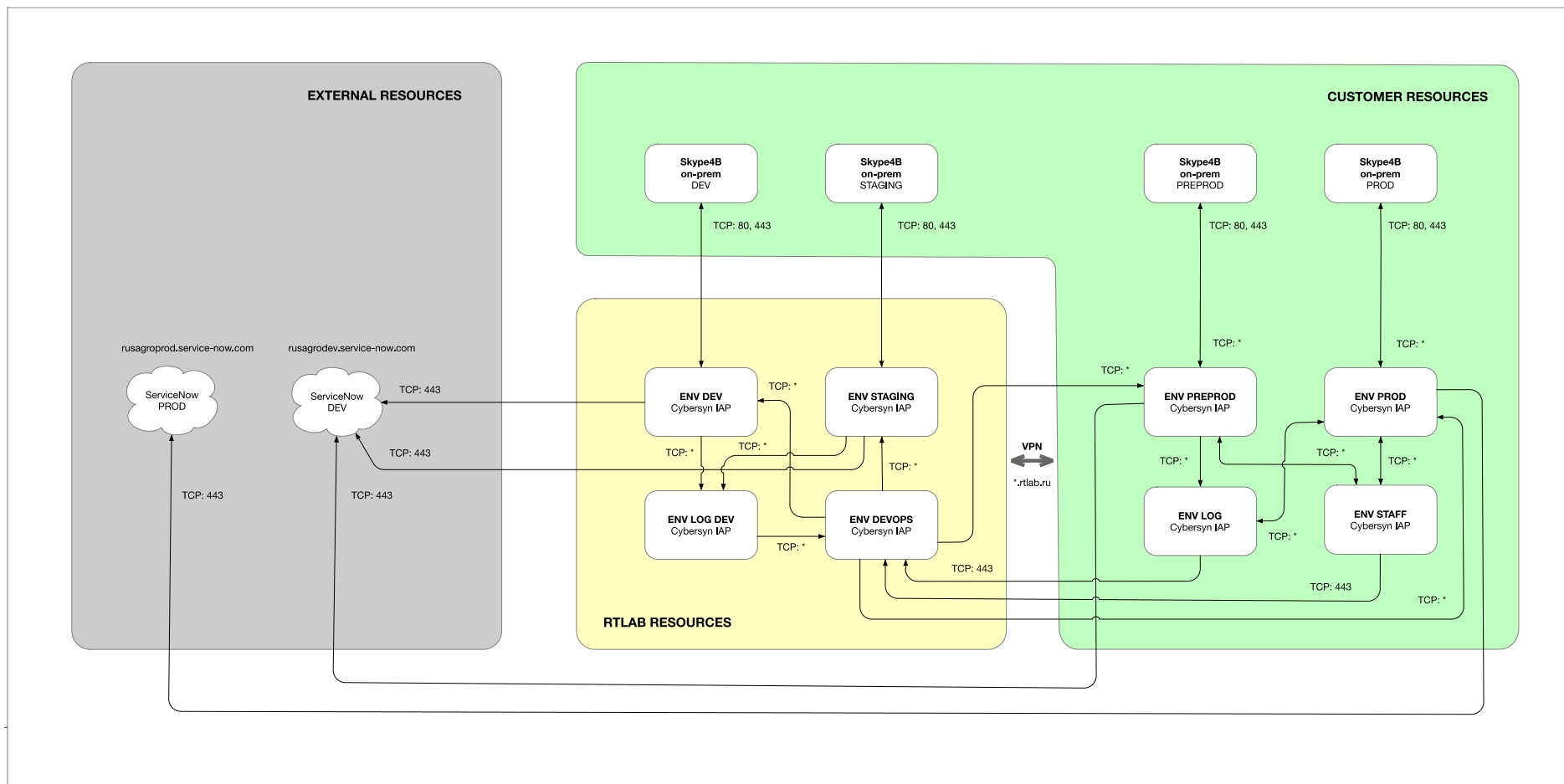


## Контуры обработки данных

- ★ KNOWLEDGE PROCESSING – контур базы знаний (актуализация и поиск знаний в автоматическом режиме)
- ★ REQUEST PROCESSING – контур обработки вопросов и ведения диалога (преобразование вопросов в ответы в автоматическом режиме)
- ★ ORDER PROCESSING – контур обработки заданий (обработка ордеров в автоматическом режиме)
- ★ MODELS PROCESSING – контур подготовки моделей (в полуавтоматическом режиме)

\*На примере проекта: конфигурация платформы в режиме виртуального помощника

# Приложение 3: Архитектура среды DevSecOps\*

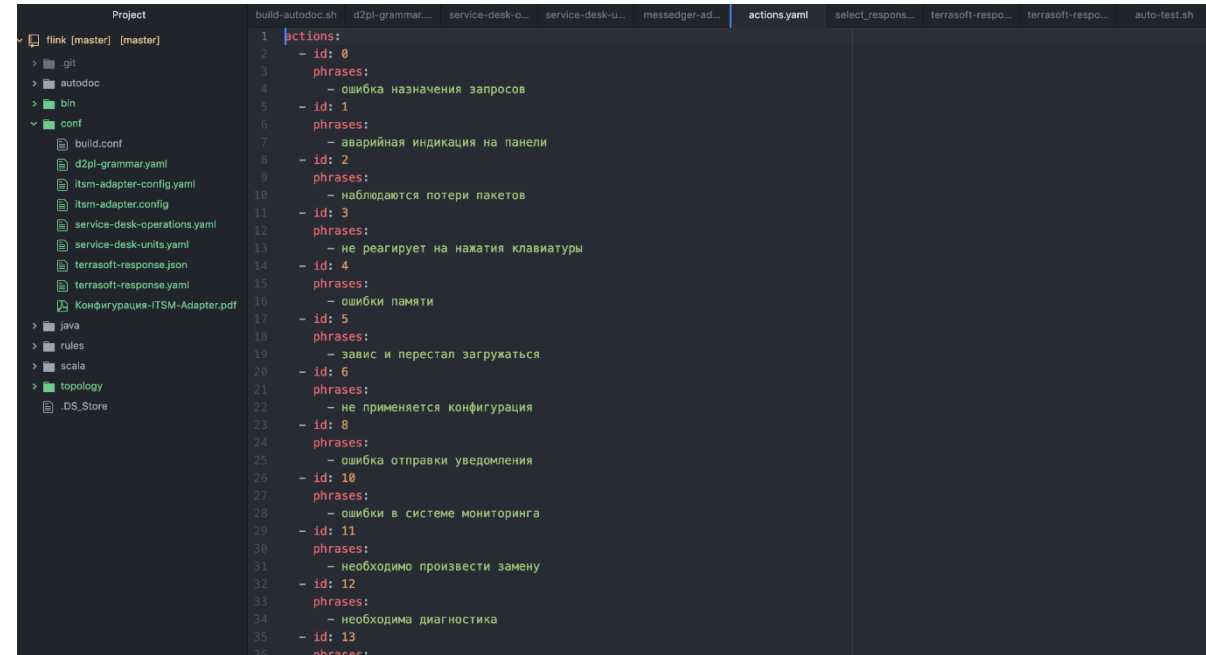


\*На примере проекта: конфигурация платформы в режиме виртуального помощника

# Приложение 5: Технологический стек виртуальной машины

- ★ OS: CentOS, Docker
- ★ Orchestration: Kubernetes\*
- ★ PL: Python, Java
- ★ DB: PostgreSQL, Cassandra, Elasticsearch, Mongo
- ★ Queueing: Kafka, ZMQ
- ★ CEP: Flink
- ★ Logging: Fluentd
- ★ Monitoring: Prometheus, Kibana
- ★ DevOps: GitLab, Redmine
- ★ REST API: Nginx, Django
- ★ ML/DL-компоненты

\*Интеграция не завершена



```
Project
├── flink [master] [master]
│   ├── git
│   ├── autodoc
│   ├── bin
│   └── conf
│       ├── build.conf
│       ├── d2pl-grammar.yaml
│       ├── itsm-adapter-config.yaml
│       ├── itsm-adapter.config
│       ├── service-desk-operations.yaml
│       ├── service-desk-units.yaml
│       ├── terrasoft-response.json
│       └── terrasoft-response.yaml
│           └── Конфигурация-ITSM-Adapter.pdf
├── java
├── rules
├── scala
├── topology
└── .DS_Store

actions.yaml
1 actions:
2   - id: 0
3     phrases:
4       - ошибка назначения запросов
5   - id: 1
6     phrases:
7       - аварийная индикация на панели
8   - id: 2
9     phrases:
10      - наблюдаются потери пакетов
11  - id: 3
12    phrases:
13      - не реагирует на нажатия клавиатуры
14  - id: 4
15    phrases:
16      - ошибки памяти
17  - id: 5
18    phrases:
19      - завис и перестал загружаться
20  - id: 6
21    phrases:
22      - не применяется конфигурация
23  - id: 8
24    phrases:
25      - ошибка отправки уведомления
26  - id: 10
27    phrases:
28      - ошибки в системе мониторинга
29  - id: 11
30    phrases:
31      - необходимо произвести замену
32  - id: 12
33    phrases:
34      - необходима диагностика
35  - id: 13
36    phrases:
```

Определение «семантических ячеек» для онтологии инцидентов

# Приложение 6: Состав библиотеки ML-компонентов

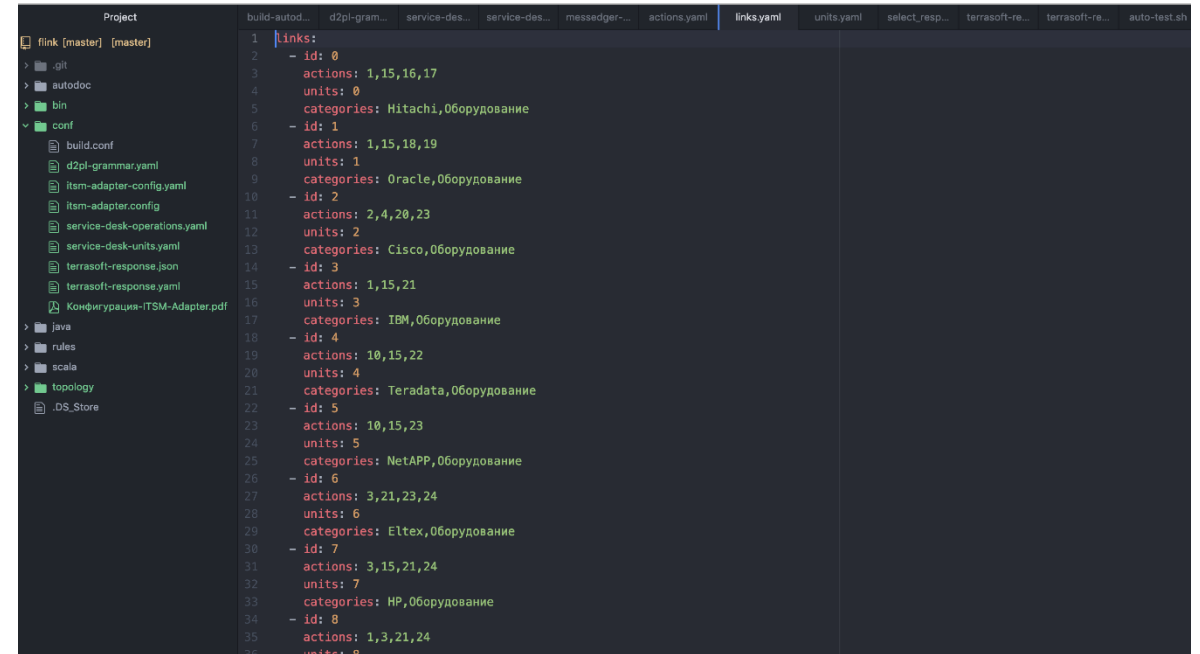
19

## ★ ML/DL-компоненты (тип «Encoder»)

- ✦ Дескриптор лица
- ✦ Код товара
- ✦ Код ошибки ПО
- ✦ Код строительного блока\*
- ✦ Код дефекта штамповки мет. деталей\*
- ✦ Код графической метки

## ★ ML/DL-компоненты (тип «Processor»)

- ✦ Идентификация лица по дескриптору
- ✦ Выдача статьи из базы знаний
- ✦ Идентификация зоны интереса
- ✦ Распознавание фото чека ККМ\*
- ✦ Локация графических меток по топологии\*
- ✦ Topic Modeling\*
- ✦ Internet User Activity Tagging\*
- ✦ Calls Tagging\*
- ✦ Calls Quality Control\*



```

Project
├── fink [master] [master]
├── .git
├── autodoc
├── bin
├── conf
│   ├── build.conf
│   ├── d2pl-grammar.yaml
│   ├── itsm-adapter-config.yaml
│   ├── itsm-adapter.config
│   ├── service-desk-operations.yaml
│   ├── service-desk-units.yaml
│   ├── terrasoft-response.json
│   ├── terrasoft-response.yaml
│   └── Конфигурация-ITSM-Adapter.pdf
├── java
├── rules
├── scala
├── topology
└── .DS_Store
links.yaml
1 links:
2 - id: 0
3   actions: 1,15,16,17
4   units: 0
5   categories: Hitachi,Оборудование
6 - id: 1
7   actions: 1,15,18,19
8   units: 1
9   categories: Oracle,Оборудование
10 - id: 2
11   actions: 2,4,20,23
12   units: 2
13   categories: Cisco,Оборудование
14 - id: 3
15   actions: 1,15,21
16   units: 3
17   categories: IBM,Оборудование
18 - id: 4
19   actions: 10,15,22
20   units: 4
21   categories: Teradata,Оборудование
22 - id: 5
23   actions: 10,15,23
24   units: 5
25   categories: NetAPP,Оборудование
26 - id: 6
27   actions: 3,21,23,24
28   units: 6
29   categories: Eltex,Оборудование
30 - id: 7
31   actions: 3,15,21,24
32   units: 7
33   categories: HP,Оборудование
34 - id: 8
35   actions: 1,3,21,24
36   units: 8
  
```

Связывание «семантических ячеек» онтологии инцидентов с «конфигурационными единицами» и «категориями каталога услуг»

\*Интеграция или разработка не завершена

# Контактные данные

20

## Прозоров Александр Александрович

Научный сотрудник Лаборатории  
машинного интеллекта МФТИ

Директор по технологиям ООО  
«Лаборатория кибер-физических  
систем»

Email: [ap@rtlab.ru](mailto:ap@rtlab.ru), [ap@cybersyn.ch](mailto:ap@cybersyn.ch)

Mobi: +7 916 9989619

