



Санкт-Петербургский государственный университет  
Кафедра информационно-аналитических систем

# Валидатор текстов выпускных квалификационных работ

**Автор:** Макеев Владислав Дмитриевич, 21.М05-мм  
**Научный руководитель:** Чернышев Георгий Алексеевич

Санкт-Петербургский государственный университет

27 ноября 2021г.

## Оглавление

1. Введение .....	3
2. Используемые технологии .....	4



Рисунок 3 – пример дерева компетенций

- SrcSeq – стандартная модель

3. Выбор датасета .....	17
3.1. 150k Python Dataset .....	17

## Оглавление

1. Введение .....	3
2. Используемые технологии .....	4

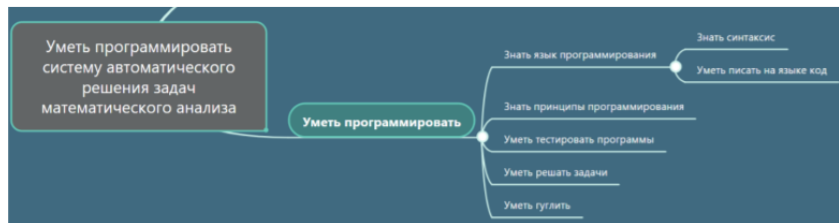
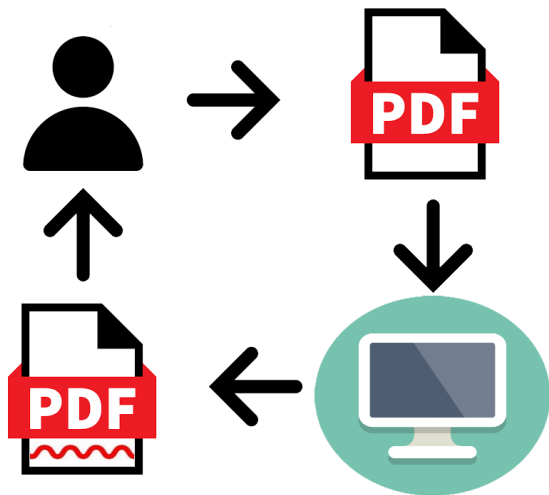


Рисунок 3 – пример дерева компетенций

- SrcSeq ☐ стандартная модель

3. Выбор датасета	17
3.1. 150k Python Dataset .....	17



# Постановка задачи

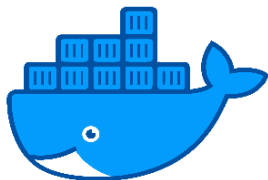
Целью работы является разработка веб-приложения, способного проверять соответствие PDF-файлов квалификационных работ определённому набору требований.

- Выбор стека технологий
- Разработка прототипа веб-приложения, позволяющего:
  - ▶ осуществлять загрузку и просмотр PDF;
  - ▶ находить и оповещать пользователя о некоторых ошибках:
    - ★ некорректное использование -, – и —;
    - ★ “сломанные” ссылки на литературу [?];
    - ★ списки из одного элемента;
  - ▶ осуществлять добавление аннотаций (подчёркивание ошибок);
  - ▶ анализировать структуру PDF (наличие оглавления, списка литературы).
- Автоматизация сборки и тестирования

- Добавление дополнительных правил
- Добавление правил с помощью высокоуровневых средств
- Подключение сторонних библиотек машинного обучения
- Подключение сторонних библиотек для обработки естественных языков

- Центрирование рисунков и подписей
- Подписи к рисункам
- Заглавные буквы в скобках
- Ссылки на литературу
- Написанное вручную оглавление
- Названия секций (капитализация, специальные символы)
- Формулы и код в виде рисунков





docker



<https://github.com/Darderion/map>

## Usage

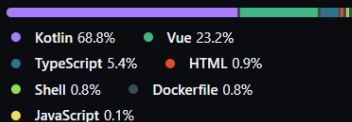
- Docker:

- `$ docker build -t map .`
- `$ docker run --rm -it -p 8080:8080 map:latest`

- Maven:

- `$ bash buildscript.sh`
- `$ cd target`
- `$ java -jar app.jar`

## Languages



- ❶ Пример для демонстрации работоспособности
- ❷ Примеры реальных работ:
  - ▶ Работа с некорректным использованием дефиса
  - ▶ Работа с некорректным использованием короткого тире
  - ▶ Работа с одним подразделом в разделе