



# СОБСТВЕННЫЙ МОДУЛЬ OCR (РАЗРАБОТКА МОДЕЛИ ПО РАСПОЗНАВАНИЮ И ВЫДЕЛЕНИЮ КОНТЕНТНОЙ ЧАСТИ ДОКУМЕНТА)

## ЗАДАЧА:

На вход системе подается набор однообразных документов. Необходимо построить систему, которая сможет заполнять электронную версию этих документов с высокой точностью. Требуется построить как средство разметки, так и модель распознаванию, основанную на технологиях TD, OCR

## РЕЗУЛЬТАТ:

- OCR: 96%
- Text Detection: 98%
- Document Classification: 99%
- Document Recognition: 94%

## РЕШЕНИЕ:

Создан программный комплекс, позволяющий произвести разметку и хранить эту информацию о документе в системе, распознать тип документа и для заданного типа документа провести заполнение его электронной формы



```
<elements>
  <element>
    <name>head</name>
    <type>Complex</type>
    <value>
      <element>
        <name>number</name>
        <type>Integer</type>
        <value>4201</value>
      </element>
      <element>
        <name>operator</name>
        <type>String</type>
        <value>Иванов</value>
      </element>
    </value>
  </element>
  <element>
  </element>
  <element>
  </element>
```

```
control-block</name>
plex</type>
int>
</element>
int>
</element>
int>
name:control-id-2</name>
type:String</type>
value>084432</value>
nent>
```

**ВИДЕО РАБОТЫ  
АЛГОРИТМА**