

Описание задачи

Написать скрипт для генерации html страницы со списком поставляемых документов в сборке и их описанием для клиента. У каждого клиента кастомная сборка, потому что у всех разный пакет документов. Список документов для данной сборки указывается в специальном bundle файле.

Описание документов должно содержаться в следующей таблице

Document type	Document description	PRADO link	String fields	Image fields	Templates

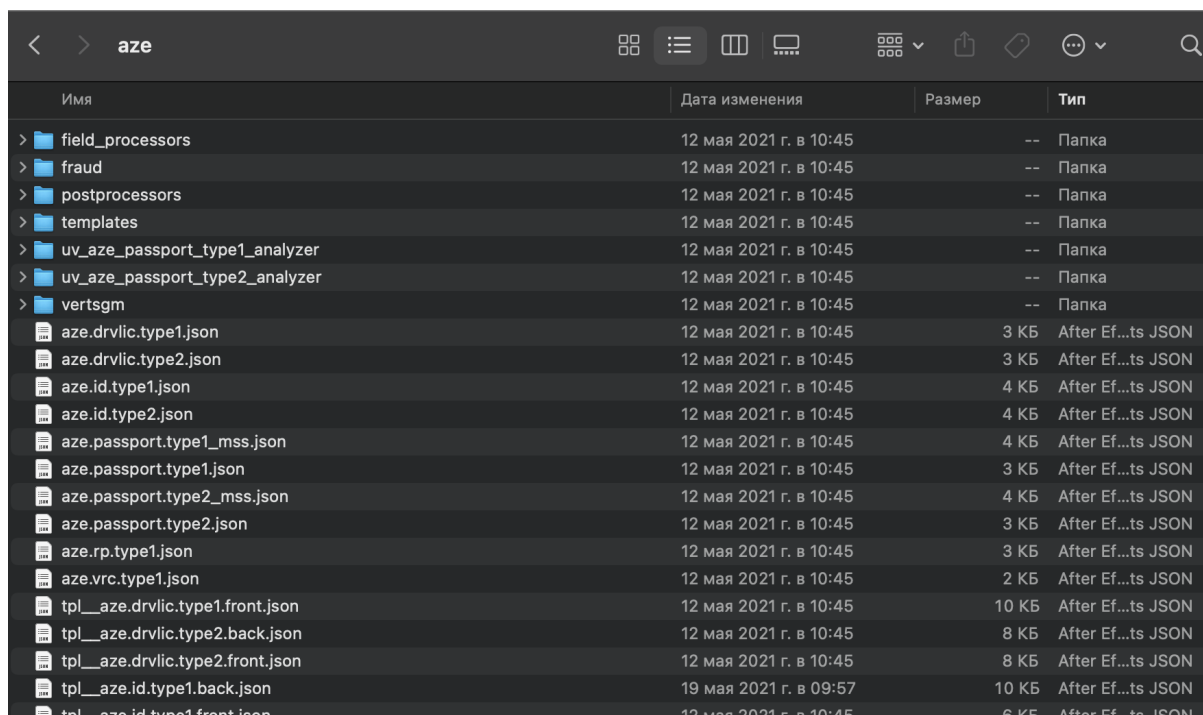
где:

- Document type — название документа. Состоит из: код страны по стандарту ISO 3166-1, вид документа (паспорт, водительское удостоверение, полис и т.д.), подвид документа (например, существует несколько подвидов водительского удостоверения: старые и новый формат). Так, aze.passport.type1 - это 1 тип паспорта Азербайджана.
- Document description — описание документа (отличительные особенности, штампы и пр).
- PRADO link — ссылка на документ в системе PRADO.
- String fields — строковые поля в документе, к которым относится фамилия, имя, дата рождения и т.д.
- Image fields — поля в виде картинок (фотография человека, подпись)
- Templates — шаблоны документа (например, у водительских удостоверений РФ две стороны, поэтому будет два шаблона).

Также необходимо сгруппировать документы по странам и сделать навигацию. Необходимо получить полное наименование страны на основе трехбуквенного кода стандарта ISO 3166-1. Требуемую информацию необходимо доставать из файлов проекта.

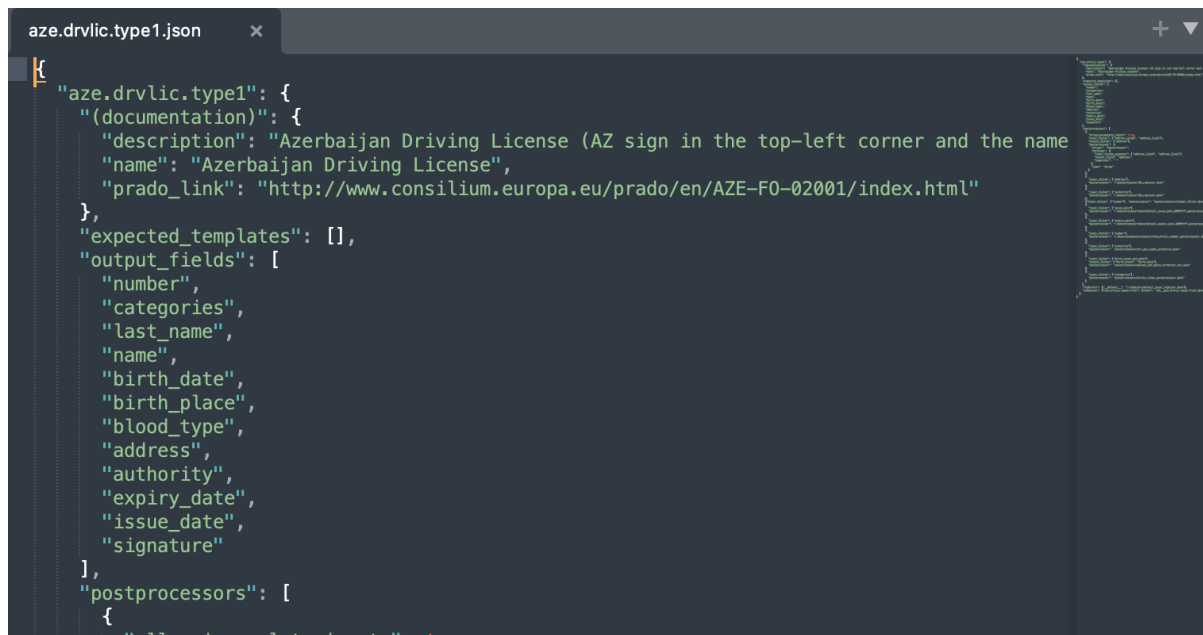
Структура проекта

Директории с документами по странам, внутри такой одной директории разные файлы для распознавания разных типов документов. На примере ниже представлены документы Азербайджана.



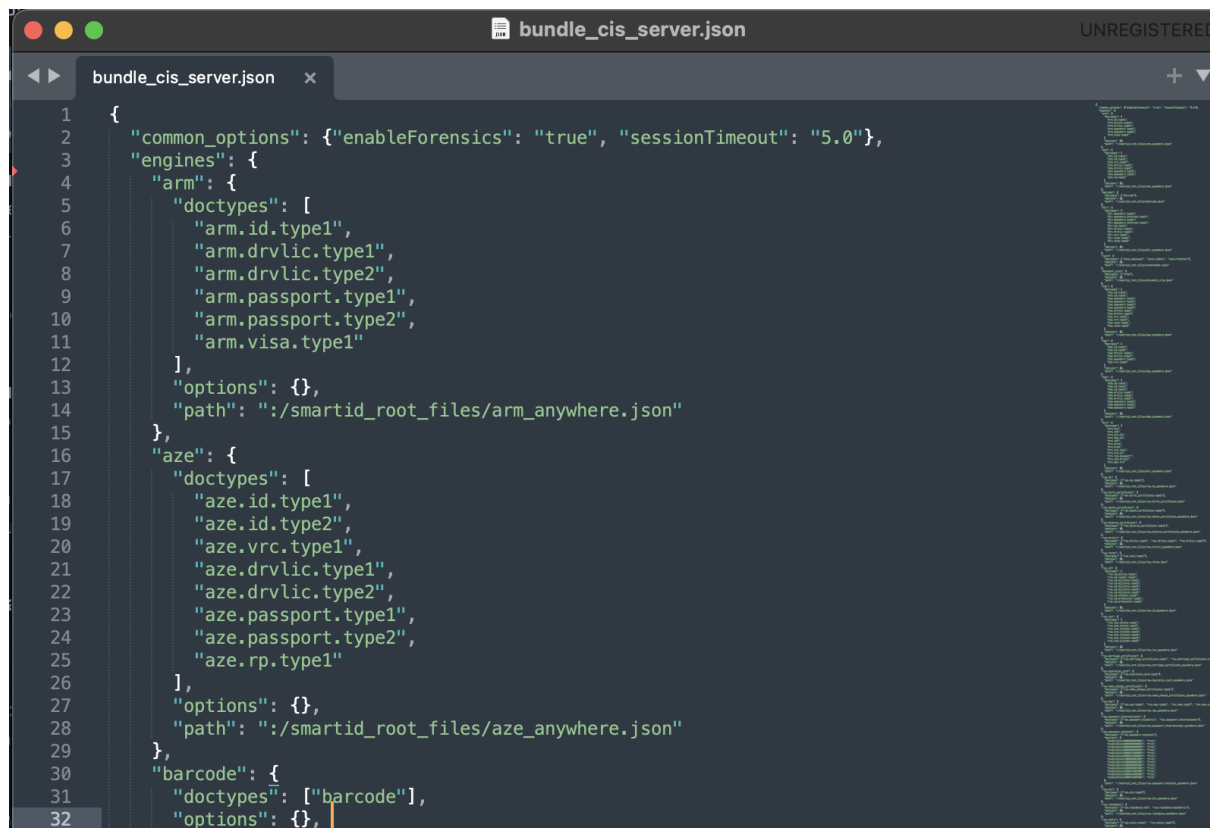
Имя	Дата изменения	Размер	Тип
> field_processors	12 мая 2021 г. в 10:45	--	Папка
> fraud	12 мая 2021 г. в 10:45	--	Папка
> postprocessors	12 мая 2021 г. в 10:45	--	Папка
> templates	12 мая 2021 г. в 10:45	--	Папка
> uv_aze_passport_type1_analyzer	12 мая 2021 г. в 10:45	--	Папка
> uv_aze_passport_type2_analyzer	12 мая 2021 г. в 10:45	--	Папка
> vertsgm	12 мая 2021 г. в 10:45	--	Папка
aze.drvlic.type1.json	12 мая 2021 г. в 10:45	3 КБ	After Ef...ts JSON
aze.drvlic.type2.json	12 мая 2021 г. в 10:45	3 КБ	After Ef...ts JSON
aze.id.type1.json	12 мая 2021 г. в 10:45	4 КБ	After Ef...ts JSON
aze.id.type2.json	12 мая 2021 г. в 10:45	4 КБ	After Ef...ts JSON
aze.passport.type1_mss.json	12 мая 2021 г. в 10:45	4 КБ	After Ef...ts JSON
aze.passport.type1.json	12 мая 2021 г. в 10:45	3 КБ	After Ef...ts JSON
aze.passport.type2_mss.json	12 мая 2021 г. в 10:45	4 КБ	After Ef...ts JSON
aze.passport.type2.json	12 мая 2021 г. в 10:45	3 КБ	After Ef...ts JSON
aze.rp.type1.json	12 мая 2021 г. в 10:45	3 КБ	After Ef...ts JSON
aze.vrc.type1.json	12 мая 2021 г. в 10:45	2 КБ	After Ef...ts JSON
tpl_aze.drvlic.type1.front.json	12 мая 2021 г. в 10:45	10 КБ	After Ef...ts JSON
tpl_aze.drvlic.type2.back.json	12 мая 2021 г. в 10:45	8 КБ	After Ef...ts JSON
tpl_aze.drvlic.type2.front.json	12 мая 2021 г. в 10:45	8 КБ	After Ef...ts JSON
tpl_aze.id.type1.back.json	19 мая 2021 г. в 09:57	10 КБ	After Ef...ts JSON
tpl_aze.id.type1.front.json	12 мая 2021 г. в 10:45	6 КБ	After Ef...ts JSON

В файле каждого документа есть документация по нему, описание полей и разные настройки распознавания:



```
{
  "aze.drvlic.type1": {
    "(documentation)": {
      "description": "Azerbaijan Driving License (AZ sign in the top-left corner and the name",
      "name": "Azerbaijan Driving License",
      "prado_link": "http://www.consilium.europa.eu/prado/en/AZE-F0-02001/index.html"
    },
    "expected_templates": [],
    "output_fields": [
      "number",
      "categories",
      "last_name",
      "name",
      "birth_date",
      "birth_place",
      "blood_type",
      "address",
      "authority",
      "expiry_date",
      "issue_date",
      "signature"
    ],
    "postprocessors": [
      {
        "allow_incomplete_inputs": true
      }
    ]
  }
}
```

Bundle файл, в котором прописаны документы требуемых стран для распознавания в требуемой сборке:



```
1  {
2    "common_options": {"enableForensics": "true", "sessionTimeout": "5.0"},
3    "engines": {
4      "arm": {
5        "doctypeypes": [
6          "arm.id.type1",
7          "arm.drvlic.type1",
8          "arm.drvlic.type2",
9          "arm.passport.type1",
10         "arm.passport.type2",
11         "arm.visa.type1"
12       ],
13       "options": {},
14       "path": ":/smartid_root_files/arm_anywhere.json"
15     },
16     "aze": {
17       "doctypeypes": [
18         "aze.id.type1",
19         "aze.id.type2",
20         "aze.vrc.type1",
21         "aze.drvlic.type1",
22         "aze.drvlic.type2",
23         "aze.passport.type1",
24         "aze.passport.type2",
25         "aze.rp.type1"
26       ],
27       "options": {},
28       "path": ":/smartid_root_files/aze_anywhere.json"
29     },
30     "barcode": {
31       "doctypeypes": ["barcode"],
32       "options": {},
```

Результат

Используемый язык: python

Подключенные модули:

- argparse — модуль для обработки аргументов командной строки
- utils.common_utils — локальный модуль, позволяет работать с json файлами в требуемом для компании формате
- os.path — модуль для работы с путями
- yattag — библиотека для генерации HTML страницы
- base64 предоставляет функции для кодирования двоичных данных в печатаемые символы ASCII и декодирования таких кодировок обратно в двоичные данные
- PIL — библиотека, предназначенная для работы с растровой графикой
- io предоставляет основные средства Python для работы с различными типами ввода-вывода
- textwrap используется для переноса и форматирования простого текста
- русcountry — библиотека, используемая для преобразования кода страны в полное название страны
- sys обеспечивает доступ к некоторым переменным и функциям, взаимодействующим с интерпретатором python
- json позволяет кодировать и декодировать данные json формата

Готовая страница:

file:///C:/repo/scripts/doc/DOC_REF.html

List of countries and special type of documents in this bundle

1 | Armenia

2 | Azerbaijan

3 | Barcode

4 | Belarus

5 | Card

6 | Kazakhstan

7 | Kyrgyzstan

8 | Moldova, Republic of

9 | Russian Federation

10 | Special mrz

11 | Tajikistan

12 | Turkmenistan

13 | Ukraine

14 | Uzbekistan

Document type	Document description	PRADO link	String fields	Image fields	Templates
Armenia					
arm.drvlic.type1	Armenian Driving License (stamp with Armenian emblem on the right, front)	PRADO link	address_eng authority_eng birth_date birth_place birth_place_eng birth_place_rus categories expiry_date id_number issue_date last_name last_name_eng last_name_rus name name_eng name_rus nationality number remarks	photo signature	arm.drvlic.type1.back arm.drvlic.type1.front
	Armenian Driving License (stamp with Armenian emblem on the		address address_eng authority authority_eng birth_date birth_place birth_place_eng categories expiry_date	photo	arm.drvlic.type2.front

file:///C:/repo/scripts/doc/DOC_REF.html#card

Top

Запуск скрипта из командной строки:

```
evampilova@DESKTOP-JGDAV41:/mnt/c/repo/scripts$ python make_documents_reference_doc.py --bundle ../data/bundle_cis_server.json --out-html ../doc/DOC_REF.html
```

Аргументы:

- bundle (путь к bundle файлу)
- out-html (путь для HTML страницы)