

## ПРИМЕНЕНИЕ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ СИМВОЛОВ ДЛЯ ОЦИФРОВКИ МАССИВНЫХ ПЕЧАТНЫХ ФОРМ

*Никулин Федор Николаевич (staff@vniief.ru), Занькова Ольга Николаевна,  
Танаева Елена Михайловна, Меньшиков Олег Владиславович*

ФГУП «РФЯЦ-ВНИИЭФ», г. Саров Нижегородской обл.

В данной работе отражены результаты работ по созданию модуля оптического распознавания массивных печатных форм карт анализа-несоответствия разрабатываемого в рамках реализации комплекса программ управления предприятием в составе комплекса программ «Система полного жизненного цикла «Цифровое предприятие» (СПЖЦ.MES v3).

**Ключевые слова:** оптическое распознавание символов, предобработка, карта анализа несоответствия, интеграция.

## APPLICATION OF OPTICAL CHARACTER RECOGNITION FOR DIGITIZATION OF MASSIVE PRINTED FORMS

*Nikulin Fyodor Nikolaevich (staff@vniief.ru), Zankova Olga Nikolaevna,  
Tanaeva Elena Mikhailovna, Menshikov Oleg Vladislavovich*

FSUE «RFNC-VNIIEF», Sarov Nizhny Novgorod region

This paper reflects the results of the work on the creation of a module for optical recognition of massive printed forms of analysis-discrepancy maps developed as part of the implementation of a set of enterprise management programs as part of the "Full Life Cycle System "Digital Enterprise"" (PLM.MES v3).

**Keywords:** optical character recognition, preprocessing, discrepancy analysis map, integration.

### Введение

Основываясь на распоряжении председателя Правительства Российской Федерации Михаила Мишустина от 6 ноября 2021 г. № 3142-р, во всех областях промышленности следует провести активное внедрение в производственную практику шести ключевых инноваций: искусственного интеллекта, робототехники, интернета вещей, технологий дополненной реальности, новых производственных и коммуникационных технологий. Они в свою очередь помогут повысить производительность труда, снизят себестоимость продукции и ускорят вывод товаров на рынок [1].

Проводящийся в данный момент процесс цифровизации производства и переноса уже имеющихся накопленных архивных данных из печатных носителей в цифровые копии повлек за собой ряд проблем.

Одной из них, в разрезе разработки системы контроля качества был перенос данных из массивных таблиц карт анализа несоответствия в «цифровой архив». Изначально, предполагалось ручное занесение всех данных из ячеек в таблицы системы человеком. Этот процесс чрезвычайно трудоемок и влечет за собой конфликты при внедрении программного обеспечения (ПО) на производстве. Решение данной проблемы возможно с использованием передовых технологий в области искусственного интеллекта. Даже частичная автоматизация переноса данных с вещественных носителей в цифровые на порядок ускоряет цифровизацию и смягчает возникающие конфликты в процессе трансформации производства. Начало работы было посвящено анализу объекта распознавания и существующих методов решения подобных проблем. Образец карты анализа-несоответствия (КАН) представлен на рис. 1 [2].





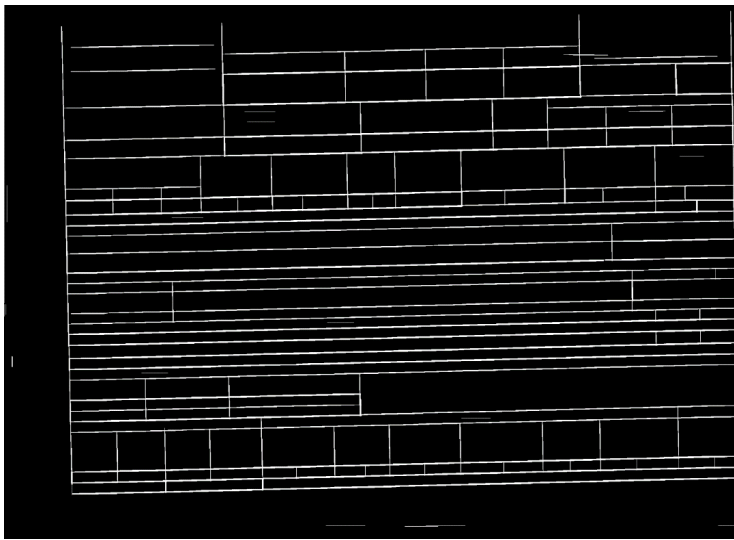


Рис. 4. КАН после порогового и морфологического преобразования лицевой стороны

КАРТА АНАЛИЗА НЕСООТВЕТВИЯ															УТВЕРЖДАЮ (2)																	
Вид документа (3)			Структурное звено (4)			Номер КАН (5)			Дата оформления (6)			Дата утверждения (7)																				
Протоколы испытаний №1 (код 13)			4			1			2021																							
Наименование изделия (8)			Обозначение изделия/обозначение изделия, в которое вводит несоответствующее изделие (9)			Заводской номер (10)			Дата изготовления (11)			Количество изделий (12)			отклонения (14)																	
Корпус			30			2			2021			22			10																	
Наработка изделия до обнаружения некорректного состояния (15)			Принадлежность изделия (16)			Предъявитель претензии (17)			Вид претензии (18)			Дата поступления КАН (19)			Вид контроля, испытаний (20)			Воздействующий фактор (21)			Влияние на работоспособность (22)											
Цели	Час	Месяц	Код	30, 4	Код	13	Код	21	Код	2021	Код	36	Код	10	Код	1	Код	359														
Характер выявленного несоответствия (23) электротехническое сопротивление изоляции не соответствует требованию КД (должно быть не менее - ЗМО), фактически значение: Ом, МОм, 1 МОм, 1-10 МОм, 5 МОм, 15 МОм, 5 МОм.															ОТК (24)		Т.Г.С															
Сравнительные измерения между парами проводников 1-2 и корпусом за № 21 Ом, 5 МОм, 5 МОм, 15 МОм, 5 МОм.																																
Сравнительные измерения между парами проводников 3-4 и корпусом за № 21 Ом, 5 МОм, 5 МОм, 15 МОм, 5 МОм.																																
Данные о несоответствующих комплектующих изделиях															Изготовитель (26)																	
Обозначение (25)																																
Деталь, сборочная единица																																
Полная продукция																																
Техническая сущность несоответствия (27) не соответствует электротехническое сопротивление изоляции изделия.																	Код (28) 452															
Причина несоответствия (29) несоответствие сверляющие изоляции требованиям КД.																	Код (30) 452															
Выводы несоответствия: 1																																
Структурное звено (31)			ФИО (32)			Должность, подпись (33)																										
Последующий документ																																
Выводы несоответствия: 1																																
Вид документа (34)			Структурное звено (35)			Номер (36)			Дата (37)			По претензии (38)			По учёту (39)			О возмещении расходов (40)			Об использовании изделий (41)			О заделе (42)			О выпущенных изделиях (43)			Предмет анализа (44)		
												Код 1			Код 0			Код 2			Код 1			Код 2			Код 1			Код 0		
Отметка ОТК (45)																																

Рис. 5. Рамка таблицы после преобразования

КАРТА АНАЛИЗА НЕСООТВЕТВИЯ															УТВЕРЖДАЮ (2)																	
Вид документа (3)			Структурное звено (4)			Номер КАН (5)			Дата оформления (6)			Дата утверждения (7)																				
Протоколы испытаний №1 (код 13)			4			1			2021																							
Наименование изделия (8)			Обозначение изделия/обозначение изделия, в которое вводит несоответствующее изделие (9)			Заводской номер (10)			Дата изготовления (11)			Количество изделий (12)			отклонения (14)																	
Корпус			30			2			2021			22			10																	
Наработка изделия до обнаружения некорректного состояния (15)			Принадлежность изделия (16)			Предъявитель претензии (17)			Вид претензии (18)			Дата поступления КАН (19)			Вид контроля, испытаний (20)			Воздействующий фактор (21)			Влияние на работоспособность (22)											
Цели	Час	Месяц	Код	30, 4	Код	13	Код	21	Код	2021	Код	36	Код	10	Код	1	Код	359														
Характер выявленного несоответствия (23) электротехническое сопротивление изоляции не соответствует требованию КД (должно быть не менее - ЗМО), фактически значение: Ом, МОм, 1 МОм, 1-10 МОм, 5 МОм, 15 МОм, 5 МОм.															ОТК (24)		Т.Г.С															
Сравнительные измерения между парами проводников 1-2 и корпусом за № 21 Ом, 5 МОм, 5 МОм, 15 МОм, 5 МОм.																																
Сравнительные измерения между парами проводников 3-4 и корпусом за № 21 Ом, 5 МОм, 5 МОм, 15 МОм, 5 МОм.																																
Данные о несоответствующих комплектующих изделиях															Изготовитель (26)																	
Обозначение (25)																																
Деталь, сборочная единица																																
Полная продукция																																
Техническая сущность несоответствия (27) не соответствует электротехническое сопротивление изоляции изделия.																	Код (28) 452															
Причина несоответствия (29) несоответствие сверляющие изоляции требованиям КД.																	Код (30) 452															
Выводы несоответствия: 1																																
Структурное звено (31)			ФИО (32)			Должность, подпись (33)																										
Последующий документ																																
Выводы несоответствия: 1																																
Вид документа (34)			Структурное звено (35)			Номер (36)			Дата (37)			По претензии (38)			По учёту (39)			О возмещении расходов (40)			Об использовании изделий (41)			О заделе (42)			О выпущенных изделиях (43)			Предмет анализа (44)		
												Код 1			Код 0			Код 2			Код 1			Код 2			Код 1			Код 0		
Отметка ОТК (45)																																

Рис. 6. КАН выровненная по вертикали

Далее обрежем лишние края, используя полученные координаты рамки. Также сделаем легкое размытие изображения по Гауссу, это будет необходимо в дальнейшем для лучшего распознавания текста (рис. 7).

Далее, используя модифицированный метод морфологической обработки, описанный выше, находим и отсекаем все контуры на изображении. В результате получаем координаты вершин всех ячеек таблицы. Используя сортировку, получаем массив, разбитый на строки и ячейки слева направо (рис. 8).

Рис. 7. КАН после обрезки и очищения

Рис. 8. КАН с выделенными и отсортированными ячейками таблицы

В левом нижнем углу в ячейки КАН, рис. 8, указан номер строки, а в правом верхнем углу – номер ячейки в строке.

### Интеграция и создание объекта КАН в системе

На финальной стадии создании собственного модуля распознавания неоднородных таблиц вырезаем из изображения ячейки и передаем в модель распознавания TESSERACT-OCR. На выходе получаем текст рассортированный по ячейкам. Ячейки всегда расположены в стандартном порядке, поэтому для ускорения обработки передаем в модель распознавания только те ячейки, данные из которых нам нужны.

Далее формируем JSON файл с распознанным текстом и передаем обратным REST запросом в MES систему в конце обработки полученного по все тому же запросу PDF файлу (рис. 9).

Конечные пользователи получают ~ на 95 % заполненную web-форму КАН (рис. 10).

Далее, внося некоторые исправления, мы получаем полностью оцифрованный документ, состоящий из объектов готовых к систематизации и анализу. Весь процесс от нажатия на кнопку «Создать из PDF-файла» до получения заполненной формы длится несколько секунд вместо дюжины минут ручного заполнения.

```
{
  "bitingmc": "Протоколы испытаний 195-12/892;. код13)",
  "entermc": "2124",
  "": "21/Н-021",
  "": "05.08.2021",
  "": "и Корпус",
  "entmc": "АС.230/ АС.200",
  "ynumber": "211901, 211902, 211907, 241908, 211922, 211914, 211915, 211921, 211922, 211923",
  "tiondate": "июнь 2021:",
  "": "22",
  "icr": "10",
  "urn": "",
  "kindmc": "30,4",
  "querymc": "13",
  "tiontoqualitykindmc": "21",
  "icard": "05.08.2021.",
  "lkindmc": "16",
  "tormc": "10",
  "orkworkdetailmc": "1",
  "redescmc": "359",
  "redesch1": "Характер выявленного несоответствия 23 электрическое сопротивление изоляции не соответствует требованию КД (должно быть",
  "redesch11": "сопротивление изоляции между парами проводников 1-2 и 3-4 зав 211901-10-15Мом,211907-5Мом,211912-1Мом,211921-10-15Мом",
  "redesch12": "сопротивление изоляции между парами проводников 1-2 и корпусом зав 211901-0,5Мом,211902 -0,15Мом,211908-2,5Мом,211912-",
  "redesch13": "сопротивление изоляции между парами проводников 3-4 и корпусом завМе 211901-0,5 Мом,211908-5Мом,211912-5 Мом,211914-15",
  "rhl": "с ооибео ОИ",
  "redescmc1": "452",
  "ausemc": "452",
  "avodmc": "Винювик несоответствия: не установлен 1711:",
  "orkcentermc": "",
  "rhl": "",
  "sionhl": "",
  "orkcentermc2": "",
  "rhl2": "",
  "sionhl2": "",
  "itingmc1": "",
  "orkcentermc1": "",
  "fclaiimmc": "",
  "epdispntmc": "ие:",
  "freturnoutmc": "у.",
  "fusedetailmc": "РА",
  "klogproductmc": "Ух",
  "frealdetailmc": "Я",
  "analysismc": "ого",
  "tofevent0": "Содержание мероприятия 47",
  "e0": "Срок выполнения. (48)",
  "urallink0": "Структурное звено (93)",
  "": "Фамилия, подпись 50 у",
  "ок \\ (51) \\",
  "tionmark0": "отчетного документа",
  "sresults [
    "Несоответствие подтвердилось, сопротивление изоляции между парами проводников не соответствует требованиям КД,",
    "Проведен анализ КД и ТД, требуется введение в ТД контроля элентрического сопротивления изоляции обоймы в сборе АС.270 для своевр",
    "изготовления."
  ]
}
```

Рис. 9. Обезличенный пример получаемого JSON файла


Реестр КАН с [ ] с 19.02.2022 по 1

Создать из pdf-файла



Щелкните для ввода фильтра...

Реестр КАН / Создание КАН

 **Выполнено**  
 Файл успешно обработан  
 Закрыть

Создание КАН, номер: б/н, состояние: Формируется

Отмена Сохранить

**Общие реквизиты**

Заполните общие характеристики КАН

Вид первичного документа	12	ИЗВ
Учетный номер первичного документа *	2124	
Дата оформления КАН *	19.02.2022	
Структурное звено, где обнаружено несоответствие *		
Обозначение изделия по КД *		
Обозначение головного изделия *		
Дата изготовления *	01.06.2021	
Заводской номер изделия		
Принадлежность изделия по виду *	30	Сборочная единица
Принадлежность по стадии жизненного цикла *	4	Опытный образец по КД с литерой "О"
Предъявитель претензии *	13	СЦ
Количество предъявленных изделий *	22	
ФИО контролера		
Дата передачи КАН председателю комиссии		
Дата рассылки копии КАН *		

**Несоответствие**

Заполните характеристики выявленного несоответствия

Вид претензии *	Начните вводить код или наименование...	
Воздействующий фактор *	10	Нормальные условия
Характер выявленного несоответствия *	359	Прочие виды отсутствия указанной в КД. ИД электрической цепи
Описание характера несоответствия *	Характер выявленного несоответствия 23 электрическое сопротивление изоляции не соответствует требованию КД (должно быть не менее 20МОм). фактические значения сопротивление изоляции между парами проводников 1-2 и 3-4 за [ ] Ом ро сопротивление изоляции между парами проводников 1-2 и корпусом за [ ] м. Е В В сопротивление изоляции между парами проводников 3-4 и корпусом за [ ] м. Е В В сопротивление [ ] -1. [ ]	

Рис. 10. Интерфейс реестра КАН на момент января 2022 года и часть результатов распознавания (некоторые данные обезличены)

Представленная разработка помогает избавиться от человеческих конфликтов при внедрении ПО на крупные производства, упрощает работу и делает ее более привычной, по опыту пользования коммерческими продуктами. Интеграция модуля распознавания позволяет ускорить и упростить систематизацию и анализ данных и, как следствие, улучшает показатели производства.

Разработанный модуль распознавания можно легко применить практически к любым формам и таблицам, используемых на производстве, таких как: карта анализа брака, карта анализа покупного изделия и др.

В дальнейшем планируется множественная обработка документов в фоновом режиме и занесение их в «Реестр КАН» со статусом распознано и возможностью доредактировать и сформировать конечную цифровую версию. Также в планах сделать конструктор распознавания, в котором будет возможность определять любую стандартизованную форму и иметь возможность быстро настраивать распознавание под конкретное производство.

1. Распоряжение о стратегическом направлении цифровой трансформации обрабатывающей промышленности [веб-сайт] URL: <http://government.ru/docs/43743/> (дата обращения: 04.10.2021)

2. СТО А 40.4482-2013. Система менеджмента качества. Управление несоответствующей продукцией корректирующие и предупреждающие действия. Порядок проведения работ.

3. Yandex Cloud: [веб-сайт] Правила тарификации. URL: <https://cloud.yandex.ru/docs/vision/pricing> (дата обращения: 15.10.2021).

4. Google Cloud [веб сайт] Pricing. URL: <https://cloud.google.com/vision/pricing> (дата обращения: 15.10.2021).

5. Tesseract User Manual [веб сайт] Documentation. URL: <https://tesseract-ocr.github.io/tessdoc/> (дата обращения: 10.11.2021).