



4. Использование программы "1С: Документооборот" в учебном процессе высшей школы Дзюбенко А.Л., Некрылов И.И. В сборнике: Новые информационные технологии в образовании: применение технологий "1С" в условиях модернизации экономики и образования. Сборник научных трудов 16-й международной научно-практической конференции. 2016. С. 184-187.

5. Заложнев А.Ю., Заложнева Л.Л., Чистов Д.В., Шуремов Е.Л. Эволюция и принципы построения информационных систем управления предприятием // Программные продукты и системы, 2014, № 2, ISSN 0236-235X, ISSN 2311-2735. С. 34-38.

6. <https://v8.1c.ru/news/24159.htm>

7. <https://www.itweek.ru/digitalization/article/detail.php?ID=204251>

Д.М. Дмитриев, М.А. Кудрина

АВТОМАТИЗИРОВАННАЯ СИСТЕМА РАСПОЗНАВАНИЯ БЛАНКОВ МАТЕРИАЛЬНОЙ ПОМОЩИ СТУДЕНТАМ

(Самарский университет)

В настоящее время в Самарском университете применяется так называемый смешанный документооборот, при котором отдельные документы дублируются как в электронном виде, так и в бумажном вариантах. Некоторые документы требуется заполнить от руки, после чего отсканировать и внести их в долговременную память компьютера с целью хранения, обработки и передачи.

Поэтому задача автоматизация распознавания и обработки бумажной документации является актуальной. В рамках данной работы была поставлена задача обработки бланков материальной помощи студентам, так как каждый месяц имеется большое количество заявлений на матпомощь, поступающих от студентов. Ручная обработка бланков является довольно трудоемкой рутинной работой и может быть автоматизирована.

Разработка автоматизированной системы распознавания

Автоматизированная система распознавания бланков материальной помощи студентам состоит из подсистем, представленных на рисунке 1.

На рисунке 2 представлен образец бланка материальной помощи, разработанный специально для автоматизированной системы, пригодный для рукописного заполнения и машинной обработки [1].

Первым этапом обработки изображения является поиск маркеров. Данный процесс многоступенчатый и включает в себя следующие подэтапы (представлены на рисунке 3).



Рисунок 1 – Структура автоматизированной системы распознавания бланков материальной помощи студентам

Образцы символов А Б В Г Д Е Ж З И К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я 0 1 2 3 4 5 6 7 8 9 D Z M

Маркеры

САМАРСКИЙ УНИВЕРСИТЕТ
Самарский государственный университет
Федеральное государственное бюджетное образовательное учреждение высшего образования «Самарский государственный университет имени академика С.П. Королёва»

Ректору Самарского университета В.Д. Богатыреву

От студента гр.: **поле № группы**

Фамилия: **Поле фамилии**

Имя: **Поле имени**

Отчество: **Поле отчества**

Согласовано:
Председатель Совета обучающихся Д.А. Ермолаев
Председатель профсоюзной организации обучающихся С.В. Заика

В приказ _____
Ректор _____
« _____ » _____ 20 ____ г.

ЗАЯВЛЕНИЕ
Прошу оказать мне материальную помощь в связи **Поле указания причины**

Дата: **Поле даты** Подпись _____

Акт обследования материального положения
Дата: **Поле даты**
Комиссия в составе (в состав комиссии группы входят три человека, утвержденные совместным решением деканата/дирекции, студенческого совета, профбюро факультета/института)
при проверке заявления установила, что студент нуждается в материальной помощи.

Профорг _____ (подпись)
Староста _____ (подпись)

Решение комиссии факультета информатики
Дата: **Поле даты**
Выдать материальную помощь студенту группы № _____ (Ф. И. О.)
в сумме **Поле суммы** рублей.

Председатель стипендиальной комиссии _____ / Яблокова Л.В. /
Председатель студенческого совета _____ / Галанов К.В. /
Староста факультета/института _____ / Бархаткин С.А. /
Председатель профсоюзного бюро _____ / Шерухаева В.В. /

Рисунок 2 – Образец бланка материальной помощи студентам



1. Преобразование в оттенки серого.
2. Преобразование по Гауссу (размытие изображения) – размер матрицы размытия вычисляется согласно исходным размерам изображения
3. Пороговая обработка изображения (яркость до 64 преобразуется в черный, яркость 64 и выше преобразуется в белый).
4. Выделение границ на изображении.
5. Морфологическое преобразование – осуществление перекрытия и до-рисовки найденных на предыдущем этапе границ.
6. Нахождение всех контуров на изображении и помещение их в массив (после всех предыдущих преобразований будут найдены только контуры всех маркеров, так как они являются зонами с наибольшей плотностью черных пикселей).
7. Нахождение центров масс всех контуров (найденный контур маркера имеет форму, максимально приближенную к окружности, что было достигнуто размытием по Гауссу и последующей пороговой обработкой. Таким образом легко определить центр квадрата) [2].



Рисунок 3 – Этапы обработки изображения бланка для нахождения маркеров

Вторым этапом после нахождения маркеров на изображении является нахождение и исправление перекоса отсканированного листа. Для этого для найденных крайних маркеров, а точнее их центров масс вычисляется тангенс угла. После вычисления углов по всем возможным вариантам расположений маркеров друг относительно друга значение усредняется, промахи при вычислении тангенса наклона листа исключаются по критерию Шовене. Пятый маркер, находящийся на листе является ориентационным (в случае, если лист был загружен в альбомном варианте или был перевернут).



Третьим этапом является обрезка изображения и сведение его к указанным размерам ширины и высоты (пропорции ширины и высоты изображения при этом сохраняются) [3].

После преобразования изображения оно подается в подсистему разметки обработанного изображения. В данной подсистеме выполняется поиск областей изображений с полями, в которых написана информация, необходимая для получения результата. Для разметки документа используется структурный подход описания, реализованный на языке XML. Каждое поле в бланке имеет фиксированную длину (количество ячеек), название и расположение относительно начала листа (примем за начало координат верхний левый угол). Также поля могут иметь фиксированные типы данных (дату, время, строки с фиксированным форматом, целые и дробные числа и т.д.). На данном этапе формируется массив с изображениями полей, с которых будет производиться распознавание текста.

Подсистема распознавания текста представлена программой Tesseract OCR.

После распознавания текста с массива полей, на основе XML-разметки исходного изображения строится XML-документ с результатом распознавания, который может быть применен в дальнейшем для создания протоколов или иных действий по ведению информации (занесение в базу данных и т.д.) [4].

После обработки всех поступивших в систему бланков на основе полученных данных составляется протокол заседания комиссии факультета информатики по распределению материальной помощи в формате DOCX, который может быть отредактирован членами комиссии в случае необходимости.

Литература

- 1 Бланк [Электронный ресурс] // Википедия: электронная энциклопедия. 2001-2021. URL: <https://ru.wikipedia.org/wiki/Бланк> (дата обращения: 17.04.2021).
- 2 OpenCV Tutorials [Электронный ресурс] // OpenCV 2006-2021. URL: https://docs.opencv.org/master/d9/df8/tutorial_root.html (дата обращения: 17.04.2021).
- 3 Сегментация (обработка изображений) [Электронный ресурс] // Википедия: электронная энциклопедия. 2001-2021. URL: [https://ru.wikipedia.org/wiki/Сегментация_\(обработка_изображений\)](https://ru.wikipedia.org/wiki/Сегментация_(обработка_изображений)) (дата обращения: 17.04.2021).
- 4 XML [Электронный ресурс] // Википедия: электронная энциклопедия 2001-2021 URL: <https://ru.wikipedia.org/wiki/XML> (дата обращения: 17.04.2021).