УДК 004

Исследование систем онлайн распознавателей текста

Размахнина Анна Николаевна
Приамурский государственный университет имени Шолом-Алейхема
Студент

Баженов Руслан Иванович
Приамурский государственный университет имени Шолом-Алейхема
к.п.н., доцент, зав. кафедрой информационных систем, математики и методик обучения

Аннотация
На сегодняшний день существует множество онлайн сервисов для распознавания текста. В данной статье предоставлен обзор семи систем онлайн распознавателей текста, отобранных по качеству перевода текстовых документов.
Ключевые слова: распознавание текста, сканирование.

The study of systems of online resolver’s text

Razmahnina Anna Nikolaevna
Sholom-Aleichem Priamursky State University
Student

Bazhenov Ruslan Ivanovich
Sholom-Aleichem Priamursky State University
Candidate of pedagogical sciences, associate professor, Head of the Department of Information Systems, Mathematics and teaching methods

Abstract
Today there are many software tools for text recognition. This article provides an overview of the four systems of online resolver’s text, selected for the quality of translation of text documents.
Keywords: OCR, scanning.

С каждым днем все чаще возникает необходимость перевода в электронный вид документов, напечатанных на бумаге. В таком случае можно набрать текст документа вручную, что не всегда удобно и довольно трудно, или же можно использовать сканер.

Сканированием называют перевод документа с бумаги в электронный формат, в результате чего создается образ бумажного изображения. В основе любого сканера лежит один и тот же принцип. Бумага освещается световым

лучом, свет отражается и воспринимается светочувствительным элементом. Элемент переводится как набор цветных или серых точек.

В итоге сканирования создается графический файл, представляющий собой растровое изображение отсканированного документа. Растровое изображение, в свою очередь состоит из точек. Количество точек зависит от разрешения сканера и размера исходного изображения.

Таким образом, после того как документ обработан сканером, мы получаем графическое изображение сканируемого документа. Но полученный графический образ не является необходимым нам текстовым документом.

Тема распознавания текста из формата точечного графического изображения является весьма сложной и актуальной. Перевод сканируемого изображения в текстовый формат осуществляется с помощью специальных программ распознавания образов.

Весомый прорыв в данной сфере произошел лишь в последние несколько лет. До этого перевести изображение в текст, возможно было лишь, при помощи оптического распознавания и использования специально разработанных шрифтов. Такие системы назывались OCR (Optical Character Recognition — оптическое распознавание символов) [1].

Минус таких систем в том, что при любом отклонении от заданного шрифта, программы подобного рода давали сбои.

Прогресс не стоит на месте и современное развитие технологий в области распознавания образов перевернуло представление об оптическом распознавании текста. Современные системы распознаватели текста могут переводить самые сложные шрифты, а многие распознают и рукописный текст.

Цель: исследовать зависимость качества распознавания текста от разрешения сканирования для разных онлайн-систем, дать оценку качества полученного текста, составить рейтинг систем, дать рекомендации пользователю.

Проблема распознавания отсканированных документов достаточно велика и неудивительно, что создано немалое количество программ, для этой цели.

Тема распознавания текстов достаточно актуальна и вызывает интерес у большого круга пользователей.

Рассмотрим работы, посвященные системам онлайн распознавания текста.

Статья преподавателей Пермского государственного педагогического университета, освещает проблемы разработки систем распознавания рукописных и старопечатных текстов. Также предоставлен проект создания такого комплекса на основе искусственных нейронных сетей, описаны возможности системы, основные компоненты и используемые технологические решения [2].

А.Ф.Хестанова и М.В.Васильева в своей работе описывают сферы применения систем распознавания, историю их развития, рассказывают об

особенностях изображений страниц и основных этапах их предобработки для дальнейшего распознавания [3].

Профессорами и студентами Томского политехнического университета разработан и представлен метод для анализа и классификации символов на основе применения вейвлет-преобразования [4].

И.К.Двоеглазов в своей работе раскрыл вопросы распознавания рукописного текста [5].

Тханг Ла Суан рассматривает методы распознавания рукописных текстов в системах автоматизации документооборота на промышленных предприятиях [6].

Статья Д.С. Лазарева и А.А. Ненашева содержит обзор словарей в системах распознавания рукописного текста [7].

В статье А.А. Мозговой описаны результаты по внедрению элементов системы поддержки принятия решений, в графический интерфейс программы распознавания сканированного рукописного текста [8].

Так же были рассмотрены англоязычные статьи, освещающие тему оптического распознавания отсканированного текста.

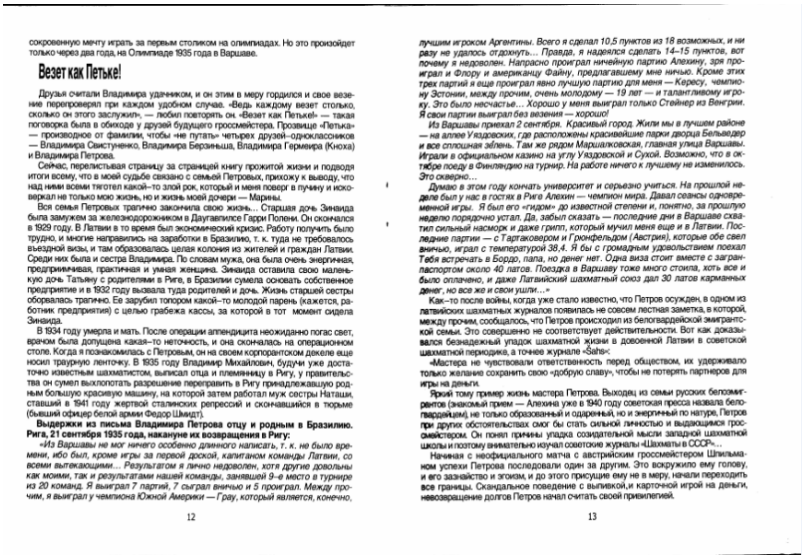
Научно-исследовательской лабораторией электроники и Департаментом электротехники и компьютерных наук в США, создана схема оптического распознавания символов «Feature selection for low error rate OCR», с целью сокращения ошибок при распознавании [9].

Sheng He, Lambert Schomaker, описывают проблему распознавания рукописного текста и предлагают метод совместного распределения, для распознавания [10].

Так как разные программы-распознаватели используют разные алгоритмы, то и результаты обработки документов получаются разными. Все потому, что современные алгоритмы распознавания текста не ориентированы на конкретные детали текста. Многие из них способны распознавать текст на разных языках, они не настроены на определенный шрифт или алфавит. Что чаще всего снижает качество перевода.

В данной статье мы рассмотрим программы, предназначенные для распознавания текста, напечатанного на русском языке.

Для исследования был взят обычный книжный текст (рис. 1) [18].



сокровенную мечту играть за первым столиком на олимпиадах. Но это произойдет только через два года, на Олимпиаде 1935 года в Варшаве.

Везет как Петьке!

Друзья считали Владимира удачником, и он этим в меру гордился и свое везение перепроверял при каждом удобном случае. «Ведь каждому везет столько, сколько он этого заслужил», — любил повторять он. «Везет как Петьке!» — такая поговорка была в обиходе у друзей будущего гроссмейстера. Прозвище «Петька» — производное от фамилии, чтобы «не путать» четырех друзей-одноклассников — Владимира Свистученко, Владимира Березняка, Владимира Гермеира (Кнока) и Владимира Петрова.

Сейчас, перелистывая страницу за страницей книгу прожитой жизни и подводя итоги всему, что в моей судьбе связано с семьей Петровых, прихожу к выводу, что над ними всеми тяготел какой-то злой рок, который и меня повергал в пучину и исковеркал не только мою жизнь, но и жизнь моей дочери — Марины.

Вся семья Петровых трагично закончила свою жизнь... Старшая дочь Зинаида была замужем за железнодорожником в Даугавпилсе Гарри Полени. Он скончался в 1929 году. В Латвии в то время был экономический кризис. Работу получить было трудно, и многие направились на заработки в Бразилию, т. к. туда не требовалось въездной визы, и там образовалась целая колония из жителей и граждан Латвии. Среди них была и сестра Владимира. По словам мужа, она была очень энергичная, предприимчивая, практичная и умная женщина. Зинаида оставила свою маленькую дочь Татьяну с родителями в Риге, в Бразилии сумела основать собственное предприятие и в 1932 году вызвала туда родителей и дочь. Жизнь старшей сестры оборвалась трагично. Ее зарубил топором какой- то молодой парень (кажется, работник предприятия) с целью грабежа кассы, за которой в тот момент сидела Зинаида.

В 1934 году умерла и мать. После операции аппендицита неожиданно погас свет, врачом была допущена какая-то неточность, и она скончалась на операционном столе. Когда я познакомилась с Петровым, он на своем корпорантском декале еще носил траурную ленточку. В 1935 году Владимир Михайлович, будучи уже достаточно известным шахматистом, выписал отца и племянницу в Ригу, у правительства он сумел выхлопотать разрешение перепрвить в Ригу принадлежащую родным большую красивую машину, на которой затем работал муж сестры Наташи, ставший в 1941 году жертвой сталинских репрессий и скончавшийся в тюрьме (бывший офицер белой армии Федор Шмидт).

Выдержки из письма Владимира Петрова отцу и родным в Бразилию. Рига, 21 сентября 1935 года, накануне их возвращения в Ригу:

«Из Варшавы не мог ничего особенно длинного написать, т. к. не было времени, ибо был, кроме игры за первой доской, капитаном команды Латвии, со всеми вытекающими... Результатом я лично недоволен, хотя другие довольны как моими, так и результатами нашей команды, занявшей 9-е место в турнире из 20 команд. Я выиграл 7 партий, 7 сыграл вничью и 5 проиграл. Между прочим, я выиграл у чемпиона Южной Америки — Грау, который является, конечно,

12

лучшим игроком Аргентины. Всего я сделал 10,5 пунктов из 18 возможных, и ни разу не удалось отдохнуть... Правда, я надеялся сделать 14–15 пунктов, вот почему я недоволен. Напрасно проиграл ничейную партию Алехину, зря проиграл и Флору и американцу Файну, предлагавшему мне ничью. Кроме этих трех партий я еще проиграл явно лучшую партию для меня — Кересу, чемпиону Эстонии, между прочим, очень молодому — 19 лет — и талантливому игроку. Это было несчастье... Хорошо у меня выиграл только Стейнер из Венгрии. Я свои партии выиграл без везения — хорошо!

Из Варшавы приехал 2 сентября. Красивый город. Жили мы в лучшем районе — на аллее Уяздовских, где расположены красивейшие пажи дворца Бельведер и все сплошная зелень. Там же рядом Маршалковская, главная улица Варшавы. Играли в официальном казино на углу Уяздовской и Сухой. Возможно, что в октябре поеду в Финляндию на турнир. На работе ничего к лучшему не изменилось. Это скверно...

Думаю в этом году кончать университет и серьезно учиться. На прошлой неделе был у нас в гостях в Риге Алехин — чемпион мира. Давал сеансы одновременной игры. Я был его «гидом» до известной степени и, понятно, за прошлую неделю порядочно устал. Да, забыл сказать — последние дни в Варшаве схватил сильный насморк и даже грипп, который мучил меня еще и в Латвии. Последние партии — с Тартаковером и Гронфельдом (Австрия), которые обе свел вничью, играл с температурой 38,4. Я бы с громадным удовольствием поехал Тебя встречать в Бордо, папа, но денег нет. Одна виза стоит вместе с заграпаспортом около 40 латов. Поездка в Варшаву тоже много стоила, хоть все и было оплачено, и даже Латвийский шахматный союз дал 30 латов карманных денег, но все же и свои ушли...»

Как-то после войны, когда уже стало известно, что Петров осужден, в одном из латвийских шахматных журналов появилась не совсем лестная заметка, в которой, между прочим, сообщалось, что Петров происходил из белогвардейской эмигрантской семьи. Это совершенно не соответствует действительности. Вот как доказывался безнадежный упадок шахматной жизни в довоенной Латвии в советской шахматной периодика, а точнее журнале «Sahs»:

«Мастера не чувствовали ответственность перед обществом, их удерживало только желание сохранить свою «добрую славу», чтобы не потерять партнеров для игры на деньги.

Яркий тому пример жизнь мастера Петрова. Выходец из семьи русских белоэмигрантов (знакомый прием — Алехина уже в 1940 году советская пресса называла белогвардейцем), не только образованный и одаренный, но и энергичный по натуре, Петров при других обстоятельствах смог бы стать сильной личностью и выдающимся гроссмейстером. Он понял причины упадка созидательной мысли западной шахматной школы и поэтому внимательно изучал советские журналы «Шахматы в СССР»...

Начиная с неофициального матча с австрийским гроссмейстером Шпильманом успехи Петрова последовали один за другим. Это вскружило ему голову, и его зазнайство и эгоизм, и до этого присущие ему не в меру, начали переходить все границы. Скандальное поведение с выпивкой,и карточной игрой на деньги, невозвращение долгов Петров начал считать своей привилегией.

13

Рисунок 1. Текст для исследования

Сервис Google Drive.

Одним из популярных онлайн сервисов по распознаванию текста, является Google Диск. Сервис бесплатный и для того чтоб получить доступ к его инструментам, достаточно иметь свой аккаунт на Google [11].

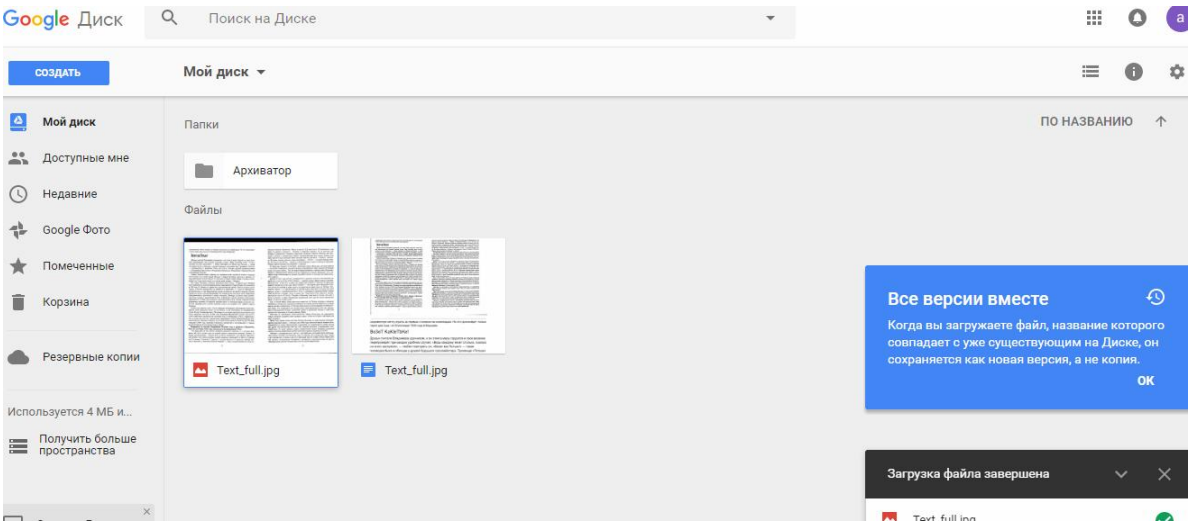


Рисунок 2. Онлайн сервис Google Диск

Алгоритм обработки графического изображения довольно простой, необходимо лишь загрузить файл на сервис.

Система оптического распознавания от Google очень быстро справилась с распознаванием.

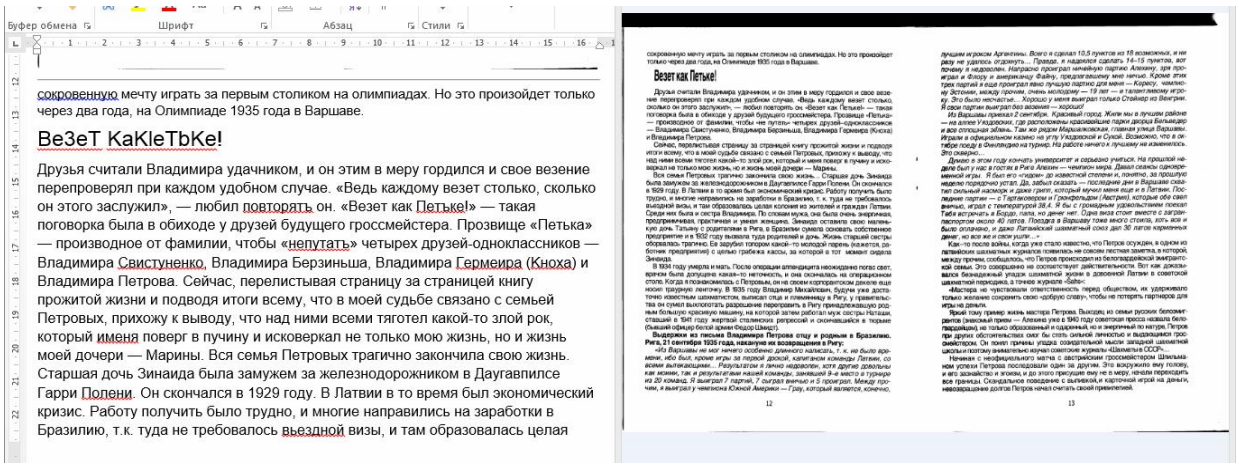


Рисунок 3. Результат распознавания сервисом GoogleДиск

Система ABBYY FineReader: Home Edition.

ABBYYFineReader – это профессиональная система оптического распознавания текста. На сегодняшний день это самая популярная система в данной отрасли, ее применяют как для домашнего использования, так и для работы в крупных организациях. Единственный минус, который можно отметить для программы – полная версия программы платная. Единственное, что следует иметь в виду – в полной версии программа платная. Однако есть возможность скачать бесплатную версию, которая рассчитана на работу в течение 15 дней и обработку 50 страниц. ABBYYFineReaderотлично поддерживает кириллицу, позволяет распознавать множество шрифтов[12].

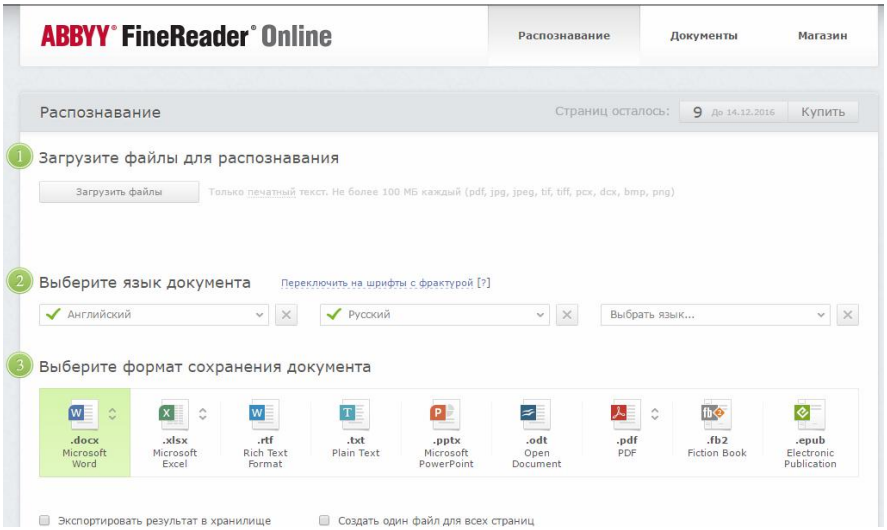


Рисунок 4. Официальный сайт система распознавания текста ABBYYFineReader

Для эксперимента мы использовали демонстрационную версию программы.

Можно отметить, что программа отлично справилась со своей задачей. Практически без ошибок. Кроме того, система автоматически создала файл в формате doc. (Рис. 3).

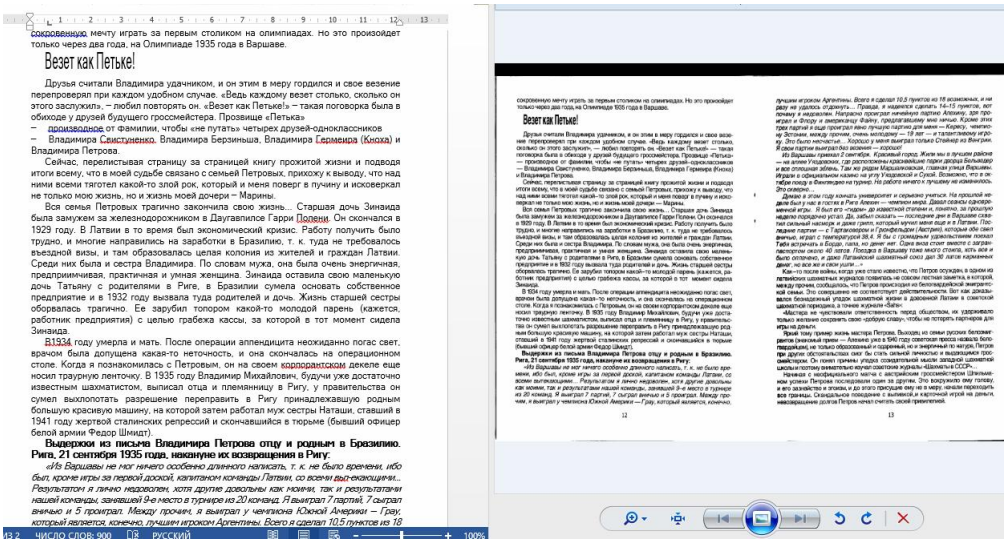


Рисунок 5. Результат распознавания программы ABBYYFineReader

Сервис Online-Ocr.

Данная система так же платная. Для доступа к распознаванию, необходимо купить «кредиты», за которые можно работать с сервисом Online-Ocr.

В демонстрационном режиме можно получить обработку лишь нескольких абзацев текста, при этом без форматирования. Файл для распознавания не должен превышать 20 Мбайт [13].

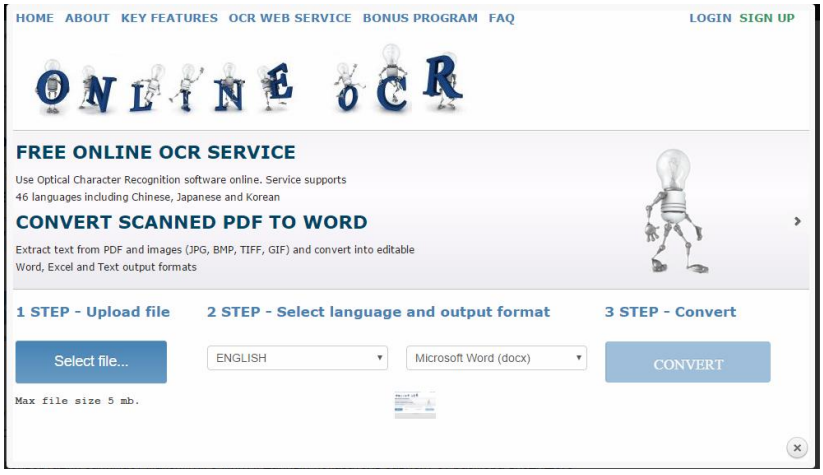
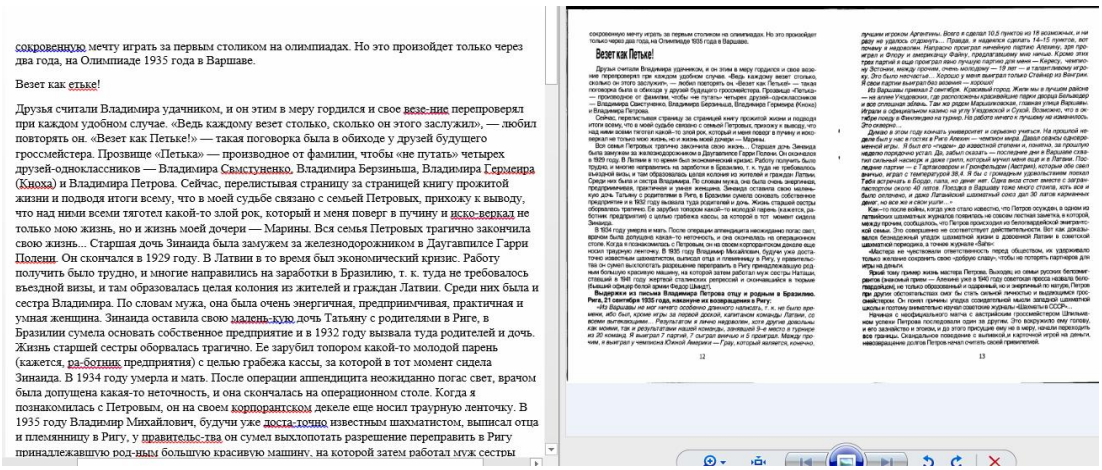


Рисунок 6. Оффициальный сайт сервиса Online-Ocr

Работа с отсканированным фрагментом текста не удалась. Качество текста оставляет желать лучшего, и исправлять опечатки в нем может быть сложнее, чем напечатать текст самому. Процесс обработки занимает не более пяти минут. Мы производили распознавание в деморежиме.



сокровенную мечту играть за первым столиком на олимпиадах. Но это произойдет только через два года, на Олимпиаде 1935 года в Варшаве.

Везет как ётьке!

Друзья считали Владимира удачником, и он этим в меру гордился и свое везе-ние перепроверял при каждом удобном случае. «Ведь каждому везет столько, сколько он этого заслужил», — любил повторять он. «Везет как Петьке!» — такая поговорка была в обиходе у друзей будущего гроссмейстера. Прозвище «Петька» — производное от фамилии, чтобы «не путать» четырех друзей-одноклассников — Владимира Свмстуненко, Владимира Берзиньша, Владимира Гермеира (Кноха) и Владимира Петрова. Сейчас, перелистывая страницу за страницей книгу прожитой жизни и подводя итоги всему, что в моей судьбе связано с семьей Петровых, прихожу к выводу, что над ними всеми тяготел какой-то злой рок, который и меня поверг в пучину и иско-веркал не только мою жизнь, но и жизнь моей дочери — Марины. Вся семья Петровых трагично закончила свою жизнь... Старшая дочь Зинаида была замужем за железнодорожником в Даугавпилсе Гарри Подени. Он скончался в 1929 году. В Латвии в то время был экономический кризис. Работу получить было трудно, и многие направились на заработки в Бразилию, т. к. туда не требовалось въездной визы, и там образовалась целая колония из жителей и граждан Латвии. Среди них была и сестра Владимира. По словам мужа, она была очень энергичная, предприимчивая, практичная и умная женщина. Зинаида оставила свою малень-кую дочь Татьяну с родителями в Риге, в Бразилии сумела основать собственное предприятие и в 1932 году вызвала туда родителей и дочь. Жизнь старшей сестры оборвалась трагично. Ее зарубил топором какой-то молодой парень (кажется, ра-ботник предприятия) с целью грабежа кассы, за которой в тот момент сидела Зинаида. В 1934 году умерла и мать. После операции аппендицита неожиданно погас свет, врачом была допущена какая-то неточность, и она скончалась на операционном столе. Когда я познакомилась с Петровым, он на своем корпорантском декеле еще носил траурную ленточку. В 1935 году Владимир Михайлович, будучи уже доста-точно известным шахматистом, выписал отца и племянницу в Ригу, у правитель-ства он сумел выхлопотать разрешение переправить в Ригу принадлежавшую род-ным большую красивую машину, на которой затем работал муж сестры

Везет как Петьке!

Рисунок 7. Результат распознавания программы Online-Ocr

Free OCR.

Система онлайн распознавания текста FreeOCR, в отличие от описанных ранее, является бесплатной. Сервис работает без регистрации, но есть ограничения на распознавание 10 документов за один час, размер которых не должен превышать 10 Мб. Данный сервис прост в использовании[14].

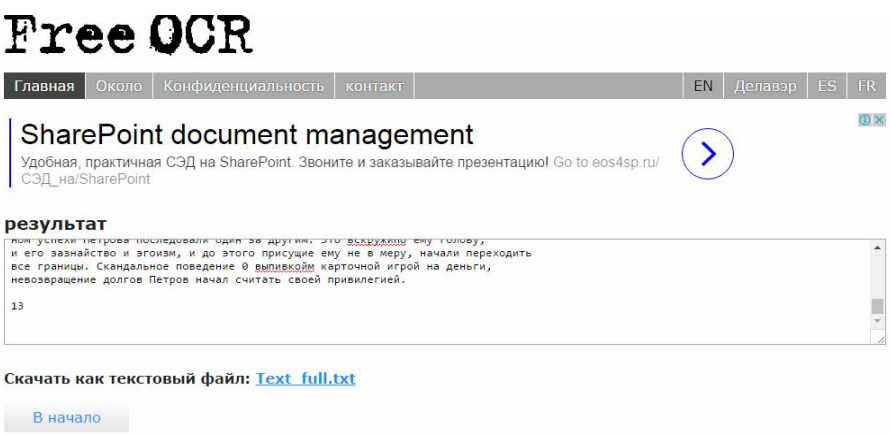
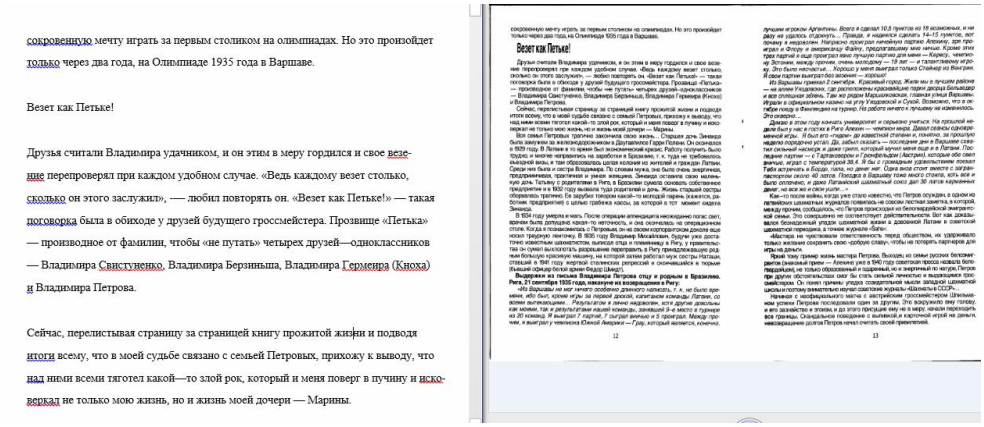


Рисунок 8. Официальный сайт программы FreeOCR

В итоге, система относительно неплохо справилась со своей задачей, не считая того, что не соблюдаются переносы и пробелы, которые пользователю придется исправлять самостоятельно.



сокровенную мечту играть за первым столиком на олимпиадах. Но это произойдет только через два года, на Олимпиаде 1935 года в Варшаве.

Везет как Петьке!

Друзья считали Владимира удачником, и он этим в меру гордился и свое везение перепроверял при каждом удобном случае. «Ведь каждому везет столько, сколько он этого заслужил», — любил повторять он. «Везет как Петьке!» — такая поговорка была в обиходе у друзей будущего гроссмейстера. Прозвище «Петька» — производное от фамилии, чтобы «не путать» четырех друзей—одноклассников — Владимира Свистуненко, Владимира Берзиньша, Владимира Гермеира (Кноха) и Владимира Петрова.

Сейчас, перелистывая страницу за страницей книгу прожитой жизни и подводя итоги всему, что в моей судьбе связано с семьей Петровых, прихожу к выводу, что над ними всеми тяготел какой—то злой рок, который и меня поверг в пучину и исковеркал не только мою жизнь, но и жизнь моей дочери — Марины.

Рисунок 9. Результат распознавания программы FreeOnlineOCR

Онлайн-сервис Free Online OCR Conversion.

Сервис Free Online OCR Conversion является бесплатным и не требует регистрации. Поддерживает форматы PDF, GIF, BMP и JPEG. Обработанный текст сохраняется в виде URL ссылки с расширением TXT. Система позволяет одновременно загружать пять файлов, с размером, не превышающим 5 Мбайт [15].

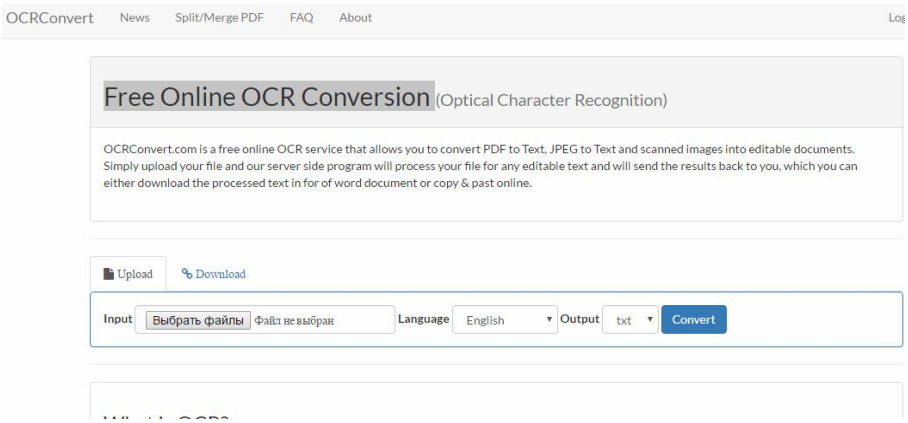
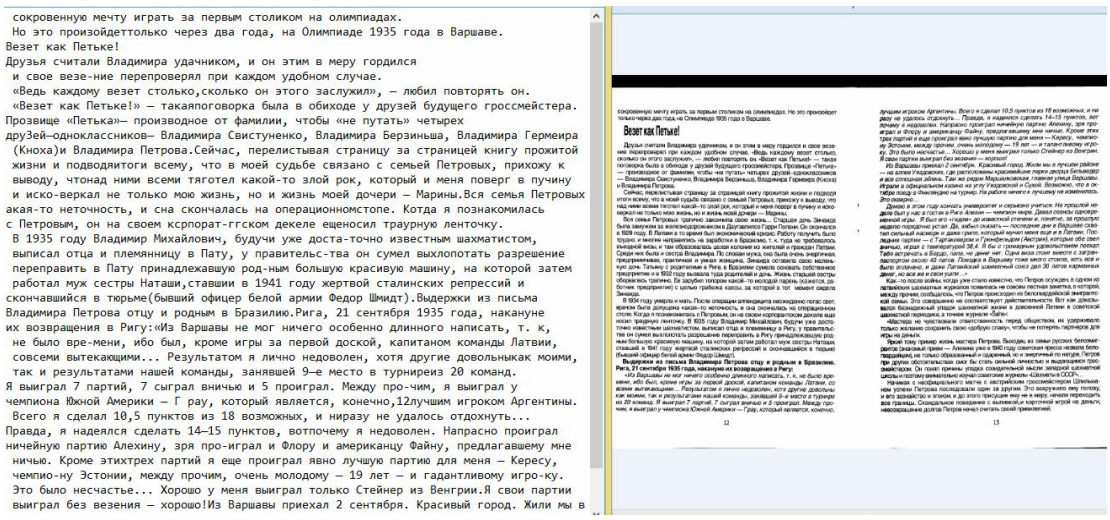


Рисунок 10. Официальная страница FreeOnlineOCRConversion

Преимуществом сервиса считается то, что он поддерживает много форматов и не требует регистрации.



сокровенную мечту играть за первым столиком на олимпиадах.
Но это произойдеттолько через два года, на Олимпиаде 1935 года в Варшаве.
Везет как Петьке!
Друзья считали Владимира удачником, и он этим в меру гордился
и свое везе-ние перепроверял при каждом удобном случае.
«Ведь каждому везет столько,сколько он этого заслужил», — любил повторять он.
«Везет как Петьке!» — такаяпоговорка была в обиходе у друзей будущего гроссмейстера.
Прозвище «Петька»— производное от фамилии, чтобы «не путать» четырех
друЗей–одноклассников– Владимира Свистуненко, Владимира Берзиньша, Владимира Гермеира
(Кноха)и Владимира Петрова.Сейчас, перелистывая страницу за страницей книгу прожитой
жизни и подводяитоги всему, что в моей судьбе связано с семьей Петровых, прихожу к
выводу, чтонад ними всеми тяготел какой-то злой рок, который и меня поверг в пучину
и иско-веркал не только мою жизнь, но и жизнь моей дочери — Марины.Вся семья Петровых
акая-то неточность, и сна скончалась на операционномстопе. Котда я познакомилась
с Петровым, он на своем ксрпорат-ггском декеле ещеносил траурную ленточку.
В 1935 году Владимир Михайлович, будучи уже доста-точно известным шахматистом,
выписал отца и племянницу в Пату, у правительс-тва он сумел выхлопотать разрешение
переправить в Пату принадлежавшую род-ным большую красивую машину, на которой затем
работал муж сестры Наташи,ставшии в 1941 году жертвой сталинских репрессий и
скончавшийся в тюрьме(бывший офицер белой армии Федор Шмидт).Выдержки из письма
Владимира Петрова отцу и родным в Бразилию.Рига, 21 сентября 1935 года, накануне
их возвращения в Ригу:«Из Варшавы не мог ничего особенно длинного написать, т. к,
не было вре-мени, ибо был, кроме игры за первой доской, капитаном команды Латвии,
совсеми вытекающими... Результатом я лично недоволен, хотя другие довольныкак моими,
так и результатами нашей команды, занявшей 9–е место в турнириез 20 команд.
Я выиграл 7 партий, 7 сыграл вничью и 5 проиграл. Между про-чим, я выиграл у
чемпиона Южной Америки — Г рау, который является, конечно,12лучшим игроком Аргентины.
Всего я сделал 10,5 пунктов из 18 возможных, и неразу не удалось отдохнуть...
Правда, я надеялся сделать 14–15 пунктов, вотпочему я недоволен. Напрасно проиграл
ничейную партию Алехину, зря про-играл и Флору и американцу Файну, предлагавшему мне
ничью. Кроме этихтрех партий я еще проиграл явно лучшую партию для меня — Кересу,
чемпио-ну Эстонии, между прочим, очень молодому — 19 лет — и гадантливому игро-ку.
Это было несчастье... Хорошо у меня выиграл только Стейнер из Венгрии.Я свои партии
выиграл без везения — хорошо!Из Варшавы приехал 2 сентября. Красивый город. Жили мы в

Везет как Петьке!

Рисунок 11. Результат распознавания программы FreeOnlineOCRConversion

После распознавания, система предлагает сохранить текст в формате TXT. С распознаванием сервис справился плохо, множество ошибок и несоблюдение переносов.

Сервис img2txt.

Img2txt – бесплатный онлайн сервис для распознавания отсканированного текста. Система обрабатывает английский, русский и украинский языки. Работает лишь с форматами изображений jpg, jpeg, png, размер которых не должен превышать 4 Мб [16].

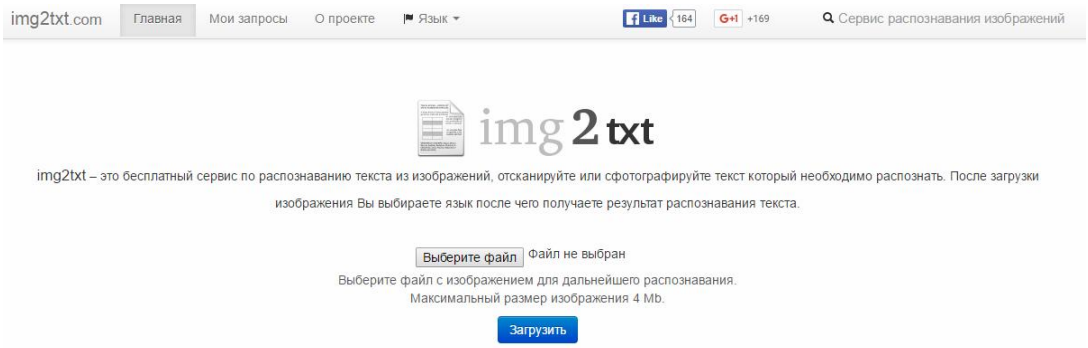
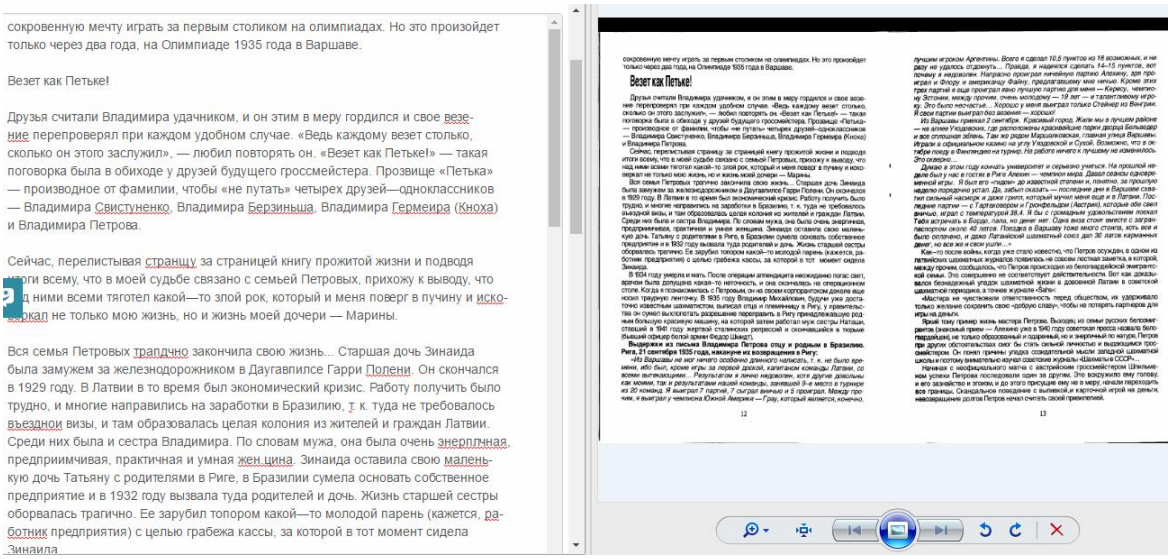


Рисунок 12. Страница сервиса img2txt

С помощью данного сервиса очень удобно распознавать отсканированный текст, но система допускает много ошибок (рис. 13).



сокровенную мечту играть за первым столиком на олимпиадах. Но это произойдет только через два года, на Олимпиаде 1935 года в Варшаве.

Везет как Петьке!

Друзья считали Владимира удачником, и он этим в меру гордился и свое везение перепроверял при каждом удобном случае. «Ведь каждому везет столько, сколько он этого заслужил», — любил повторять он. «Везет как Петьке!» — такая поговорка была в обиходе у друзей будущего гроссмейстера. Прозвище «Петька» — производное от фамилии, чтобы «не путать» четырех друзей—одноклассников — Владимира Свистуненко, Владимира Берзиньша, Владимира Гермеира (Кноха) и Владимира Петрова.

Сейчас, перелистывая страницу за страницей книгу прожитой жизни и подводя итоги всему, что в моей судьбе связано с семьей Петровых, прихожу к выводу, что над ними всеми тяготел какой—то злой рок, который и меня повергнул в пучину и исковеркал не только мою жизнь, но и жизнь моей дочери — Марины.

Вся семья Петровых трагично закончила свою жизнь... Старшая дочь Зинаида была замужем за железнодорожником в Даугавпилсе Гарри Полени. Он скончался в 1929 году. В Латвии в то время был экономический кризис. Работу получить было трудно, и многие направились на заработки в Бразилию, т. к. туда не требовалось въездной визы, и там образовалась целая колония из жителей и граждан Латвии. Среди них была и сестра Владимира. По словам мужа, она была очень энергичная, предприимчивая, практичная и умная женщина. Зинаида оставила свою маленькую дочь Татьяну с родителями в Риге, в Бразилии сумела основать собственное предприятие и в 1932 году вызвала туда родителей и дочь. Жизнь старшей сестры оборвалась трагично. Ее зарубил топором какой—то молодой парень (кажется, работник предприятия) с целью грабежа кассы, за которой в тот момент сидела Зинаида

Рисунок 13. Результат обработки программы img2txt

Онлайн сервис NewOCR.

NewOCR – бесплатный OCR сервис, поддерживающий 29 языков распознавания, включая русский. Позволяет загружать файлы в форматах JPEG, PNG, GIF, BMP, многостраничный TIFF размером до 5 Мб, а также многостраничные PDF размером до 20 Мб [17].

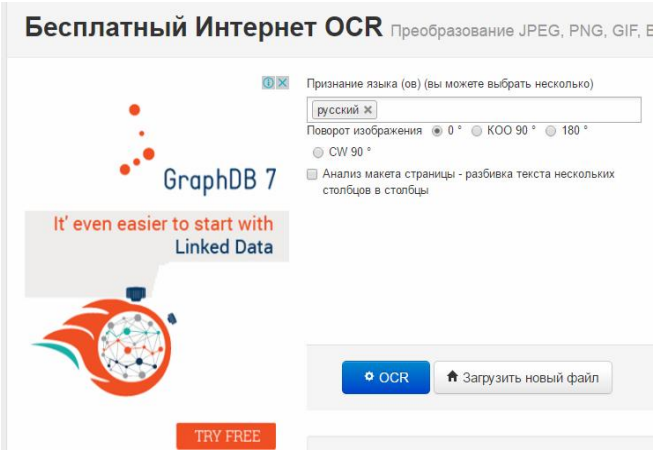


Рисунок 14. Сервис Newocr

Сервис позволяет обрабатывать изображения различных форматов. Допускает большое количество ошибок, после использования данной системы придется потратить немало времени для исправления ошибок (рис. 15).

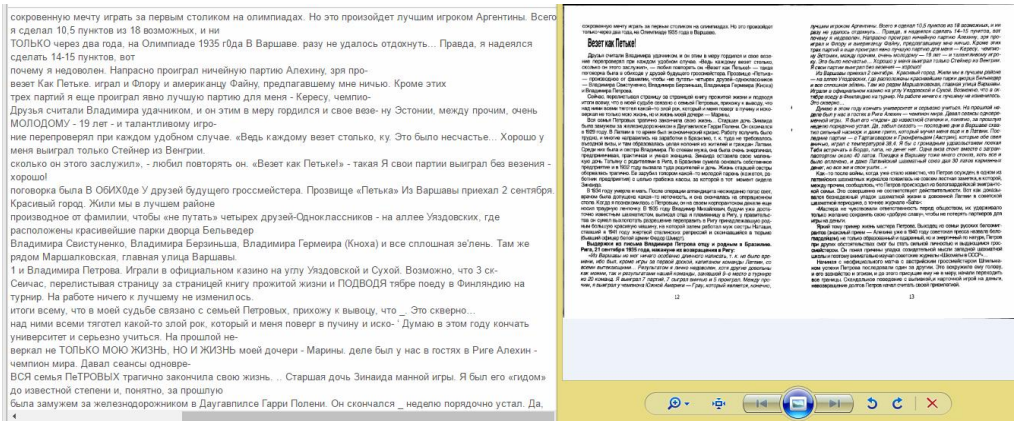


Рисунок 15. Распознавание теста системой Newocr

В ходе исследования, были рассмотрены четыре системы онлайн распознавания текста с отсканированного изображения.

Для оценки качества сервисов, сравним рассмотренные системы в таблице 1.

Таблица 1. Сравнение сервисов по онлайн-распознаванию текста

Наименование	**Количество слов с ошибкой**	**Достоинства**	**Недостатки**	**Общая оценка сервиса (1 до 10)**
Google Диск	21	- поддержка более 200 языков; - алгоритм распознавания текста очень прост;	- необходимо наличие аккаунта в Google;	4
ABBYYFineReader	2	- чтение текстов со сложным графическим оформлением; - поддержка большого числа языков; -легкость освоения;	-программа платная	9
Сервис Online-Ocr	4	- простота использования; - поддержка 46 языков;	- нет возможности сохранить текстовый файл; - полная	5

			версия программы платная;	
FreeOCR	10	- сервис бесплатный; - регистрация не нужна;	-ограничения на объем изображения;	4
FreeOnline OCRConversion	17	- сервис бесплатный;	- множество ошибок;	2
Img2txt	25	- бесплатный сервис; - простота использования;	- ограничения на размер файла; - не распознает графики, изображения, таблицы;	3
newocr	Более 50	- бесплатный сервис, не требующий регистрации;	- очень много ошибок;	2

В ходе исследования были выявлены достоинства и недостатки каждого из рассмотренных сервисов онлайн распознавания отсканированного текста. В таблице предоставлены сравнительные характеристики по каждой системе распознавания.

Сравнение производилось по количеству ошибок в слове, знаки препинания не учитывались. Оценка сервиса производилась, с учетом выявленных недостатков и ошибок, в результате распознавания.

Еще раз отметим, что были использованы программы только для онлайн распознавания.

По итогам исследования, явным лидером признан сервис ABBYYFineReader. У программы практически нет недостатков, она отлично справляется со своей задачей.

Библиографический список

1. Сканирование и распознавание // studfiles.ru URL: http://www.studfiles.ru/preview/5357061/ (дата обращения: 16.12.2016).
2. Корниенко С.И, Айдаров Ю.Р, Гагарина Д.А, Черепанов Ф.М, Ясницкий Л.Н. Программный комплекс для распознавания рукописных и старопечатных текстов // Информационные ресурсы России. 2011. №1. 35-37 с.

3. Хестанова А.Ф., Васильева М.В. Распознавание печатного текста. Основные принципы // Электронный Научный Журнал. 2016. №5(8). 153-159 с.
4. Григорьев Д.С, Хаустов П.А, Спицын В.Г. Улучшение качества метода оптического распознавания текстов с помощью совместного применения вейвлет-преобразований, курвлет-преобразований и алгоритмов словарного поиска // Известия Томского политехнического университета. Инжиниринг георесурсов. 2013. №5. 106-111 с.
5. Двоеглазов И.К. К вопросу распознавания рукописного текста // Актуальные вопросы науки и техники. 2015. 182-185 с.
6. Ла Суан Тханг. Методы распознавания рукописных текстов всистемах автоматизации документооборота на промышленных предприятиях: автореф. Дис. ... канд. техн. Наук: 05.13.06. М., 2008. 23 с.
7. Лазарев Д.С., Ненашева А.А. Использование словарей в системах распознавания рукописного текста // Решетневские чтения. 2013. №17. 223-224 с.
8. Мозговой А.А. Система поддержки принятия решений на примере распознавания сканированного рукописного текста // Вестник Воронежского государственного технического университета. 2016. №1. 25-27 с.
9. Troxel D.E. Feature selection for low error rate OCR // Pattern Recognition. 1976. №2. 73-76 с.
10. He S., Schomaker L. Beyond OCR: Multi-faceted understanding of handwritten document characteristics // Pattern Recognition. 2016. №63. 321–333 с.
11. Google диск URL: https://drive.google.com/drive/my-drive (дата обращения: 11.12.2016).
12. ABBYYFineReader. URL: https://finereaderonline.com/ru-ru (дата обращения: 06.12.2016).
13. Online-OcrURL: URL: http://www.onlineocr.net/ (дата обращения: 06.12.2016).
14. FreeOCR. URL: http://www.free-ocr.com/ (дата обращения: 06.12.2016).
15. Free Online OCR Conversion. URL: http://www.ocrconvert.com/ (датаобращения: 06.12.2016).
16. Img2txt. URL: https://img2txt.com/ (дата обращения: 11.12.2016).
17. Бесплатный Интернет OCR. Newocr. URL: http://www.newocr.com/ (дата обращения: 11.12.2016).
18. URL: http://www.3dnews.ru/571145/page-3.html (дата обращения: 11.12.2016).