

Ростовцев П.С.

*Научный руководитель: к.т.н. доцент каф. ИСПИ Салех Х.М.
Федеральное государственное бюджетное образовательное учреждение высшего образования “Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых”
600000, г. Владимир, ул. Горького, д.87
Email: p.r.ru@yandex.ru*

Анализ систем классификации документов

На сегодняшний день одной из самых решаемых задач с помощью информационных технологий, является автоматизация бизнес процессов компаний. Каждая компания ведет работу с документацией: заключает договоры, производит оплату счетов, выдает чеки и т.д. Таким образом, можно сказать работа с документами неотъемлемая часть бизнес процессов в любой компании.

Использование различных информационных технологий представляет собой конкурирующий между собой рынок. В последнее время все чаще в системы начинают внедрять нейронные сети. Мода на использование нейронных сетей в проектах распространяется не только на новые проекты, но и успешно используется для модернизации уже имеющихся решений. Решения о внедрении происходят на основаниях предоставления больших возможностей, а также поддержания конкуренции на основании мировых тенденций.

Исходя из исследований международной корпорации Microsoft, 94% предприятий считают, что технологии искусственного интеллекта могут быть использованы для решения различного рода задач их организаций. Среди опрошенных 26% предпринимателей уже внедрили данные технологии в бизнес-процессы своих компаний и еще 43% ведут разработку в данных направлениях [1].

Для того чтобы определить тенденции развития систему необходимо провести анализ уже существующих решений. В ходе анализа были выявлены четыре достаточно крупных конкурента, которые предоставляют решения по классификации документов. Также были выбраны критерии для сравнения: вид обрабатываемых документов, поддерживаемые языки, тип сервиса, назначение и стоимость.

Таблица 1 – Анализ систем классификации документов

Решения	Тип сервиса	Назначение	Вид документов	Языки	Стоимость
Smart Engine	SDK	классификация документов распознавание текста извлечение данных	паспорта водительские удостоверения штрих коды	Все языки	от 150 000 руб.
Nanonet	Веб ресурс	извлечение данных валидация данных	идентификационные карты квитанции оплаты банковские формы	Английский Китайский Арабский Португальский	от 40 000 руб. в месяц
AutoEntry	Веб ресурс	автоматический ввод данных классификация документов	финансовые документы	Английский	от 10 000 руб. в месяц
DocParser	Веб ресурс	извлечение данных классификация документов	финансовые документы банковские документы	Английский	от 15 000 руб. в месяц

В ходе анализа было выявлено что большинство решений предлагают свои услуги по обработке документов посредством веб-ресурсов. Основными видами функциями данных решений является классификация документов, и распознавание текста. Также можно отметить что чем больше возможностей имеет решение, тем выше его стоимость. Однако рассмотренные решения имеют ограниченную область применения, так как направлены на распознавание только некоторых типов документов.

Таким образом, можно сказать, что намечаются тенденции использования нейронных сетей в области документооборота. Однако существующие решения не являются многофункциональными. Использование таких средств обработки документов являются дорогостоящими. Поэтому внедрение нейронных сетей и оптимизация алгоритмов классификации может позволить усовершенствовать технологии в данном направлении и упростить работу с документами.

Литература

1. Новости. Microsoft [Электронный ресурс]. – URL: <https://news.microsoft.com/ru-ru/intelligent-economies-ais-transformation-of-industries-and-society/> (дата обращения 05.04.21)
2. Ростовцев П.С. Вершинин В.В. «Анализ актуальных направлений развития систем на основе нейросетевых технологий» стр.221-223 г. Березники, 2020 г.
3. Smart Engines: безопасное распознавание документов. [Электронный ресурс]. – URL: <https://smartengines.ru> (дата обращения 05.04.21)
4. Intelligent document processing with AI | Nanonets. [Электронный ресурс]. – URL: <https://nanonets.com> (дата обращения 05.04.21)
5. Automated data entry [Электронный ресурс]. – URL: <https://www.autoentry.com> (дата обращения 05.04.21)
6. Docparser - Document Parser Software - Extract Data From PDF [Электронный ресурс]. – URL: <https://docparser.com> (дата обращения 05.04.21)