

# Сервис для автоматической обработки писем на основе машинного обучения

*Анотация — краткое резюме проекта для быстрого ознакомления («для руководителей») объемом от четверти до половины страницы. Заинтересуйте читателя: «почему я должен тратить на это свое время?»*

Наше решение сокращает затраты на ручную обработку писем. Речь о заказах, написанных в вольном стиле: без определённой структуры и с самым разнообразным форматированием. Клиенты пишут письма так, как им удобно — а компания берёт на себя разбор этого текста и внесение заказа в базу данных. С помощью машинного обучения мы создали сервис, который выделяет из писем те данные, что нужны для заполнения формы заказа — сотрудникам это экономит около 20% рабочего времени.

*Введение — кратко об организации и ее продукции/услугах, текущее положение на момент старта проекта и исходный уровень зрелости бизнес-процессов.*

Мы организовываем туристические трансферы по всему миру — в сети более 670 аэропортов в 120 странах. Основные клиенты — агентства делового туризма и крупные корпорации. Компании заказывают поездки для сотрудников, которые отправляются в командировки.

Наша компания начала работу в 2009 году и стала пионером рынка трансферов в России. В 2011 мы запустили первую в стране систему заказа трансферов в сети — с моментальной оплатой и подтверждением заказа. Помимо сайта, система включает в себя различные инструменты для интеграций — API, виджеты и фреймы.

Мы разработали веб-сервисы и мобильные приложения для каждого из участников процесса по заказу трансферов — от пассажира и менеджера заказчика до диспетчера автопарка и водителя.

*Бизнес-контекст — вызовы, с которыми столкнулась организация, конкретные болевые точки, общие неблагоприятные тенденции или, наоборот, новые возможности, стратегические цели и новые рубежи, поставленные руководством. Располагаемые ресурсы и ключевые ограничения.*

Несмотря на развитую ИТ-инфраструктуру, многие клиенты предпочитают отправлять нам заказы не через сайт, а по почте. Менеджеру заказчика гораздо проще нажать кнопку «Переслать», чем заполнять форму заказа на нашем сайте. В этих случаях заполнять форму приходится уже нашим менеджерам — в нашей внутренней системе.

Наши менеджеры обычно переносят данные из писем в форму с помощью операции «скопировать-вставить». В среднем, для каждого заказа надо перенести порядка 10 строк:

- город;
- адреса отправления и назначения;
- время подачи автомобиля и его класс;
- имя и телефон пассажира;
- номер рейса или поезда;
- примечания.

Объём заказов растёт с каждым годом — это приводит к необходимости увеличивать и штат тех, кто обрабатывает такие письма вручную. Долгое время так и делали. Но разрастание

штата приводило к проблемам: обучение новых сотрудников обходилось дорого, часто они не справлялись с потоком писем — увеличивалось число ошибок.

В итоге пришли к задаче: исключить рост штата при увеличении объёма заказов — зафиксировать затраты на ФОТ для отдела, обрабатывающего письма.

*Бизнес-процесс — без лишних подробностей опишите целевой бизнес-процесс (бизнес-процессы), его границы и место в процессной архитектуре, участников (подразделения, роли). Кто принимал участие в проекте, кто являлся спонсором, кого проект затронул из числа сотрудников, клиентов, партнеров? Какие цели были поставлены перед процессной командой и как они связаны с бизнес-целями из предыдущего раздела?*

Было решено автоматизировать часть работы сотрудников. Ранее мы уже пробовали ускорить работу за счёт улучшения интерфейса и настройки горячих клавиш, но качественного изменения эти решения не дали. Требовалось автоматизировать сам процесс выделения нужных данных из писем.

### **Анализ решения**

В целом, для обработки писем можно было бы использовать регулярные выражения и словари. Но это помогло бы только с небольшим количеством писем — в основном от тех заказчиков, которые формируют письма автоматически в своих системах. Такое решение потребовало бы постоянной доработки условий и со временем могло бы стать слишком тяжёлым для дальнейших улучшений. Это нерентабельно.

Нужно было универсальное решение, которое позволяло бы обрабатывать письма любого вида и которое можно было бы легко масштабировать. Машинное обучение отлично подходит в качестве такого решения.

### **Ход работы**

- Для начала работы подрядчику были предоставлены данные по заказам за два последних месяца — письма в том виде как их прислали заказчики, и структурированные данные, извлечённые из этих писем.
- На первых порах подрядчику приходилось вручную размечать текст писем, чтобы обозначить связь данных из писем с тем, что было занесено в форму заказа.
- За основу архитектуры была взята нейронная сеть от Гугл — BERT. Далее её дорабатывали под решение именно нашей задачи.
- Дополнительно потребовалось ввести алгоритмы для предварительной обработке отдельных полей в соответствии с их спецификой.
- Помимо распознавания именованных сущностей (Named-Entity Recognition) для выделения отдельных частей текста также потребовалось связывание именованных сущностей (Named Entity Linking) — например, отдельную сложность представляло разделение в сущности одинаковых адресов отправления и назначения.
- Постепенно обучение модели перешло от работы с набором из двух частей (письмо/структура, выделенная из письма человеком) к набору их трёх частей: письмо/структура, выделенная из письма моделью/структура, выделенная из письма человеком.
- В процессе работы мы наладили постоянное пополнение данных для непрерывного обучения сервиса распознавания.
- Первые испытания прошли спустя три месяца: процент распознавания был низкий — 20-30%.
- Продолжая улучшать сервис, достигли качества в 40%, но далее не могли сдвинуться на протяжении нескольких месяцев.

- Потребовался глубокий анализ специфики данных в отдельных полях и доработка логики их обработки — в отдельности по каждому полю.
- Также пришлось пересмотреть систему оценки качества, которая давала некорректную оценку, сопоставляя данные один к одному. Однако при проверке вручную выяснилось, что многие моменты, оцененные машиной как ошибка, на самом деле ошибками не являются. Отличия были в отдельных символах, которые не критичны при чтении текста человеком.
- Решение дорабатывалось в течение полугода и в итоге мы дошли до 70-75% точно распознанных данных, что изначально обозначали как минимальную планку для этой системы.

*Инновационность — чем ваш проект уникален, как он позволяет вашей компании выделиться из общего ряда конкурентов? Инновационность с точки зрения бизнеса: новые бизнес-модели, новый подход к работе с клиентами, ...? Применение или разработка инновационных методологий управления бизнес-процессами. Инновационные приемы управления проектом и управления изменениями, обеспечившие успех проекта.*

Машинное обучение в нашей сфере пока никем не используется — большинство процессов реализуется через ручной труд. Но это не даёт масштабировать бизнес в короткие сроки без потери в качестве сервиса: ошибки, которые допускают сотрудники — основная причина недовольства клиентов.

Мы повышаем качество и скорость обработки данных и сокращаем риск ошибок. Наши клиенты отмечают быстроту реакции на их письма.

Но самое главное — мы позволяем клиентам передавать нам заказы в любом удобном для них виде, не неся при этом потерь на обработку данных.

*Трудности — какие уроки вы извлекли из опыта проекта? Какие советы можете дать организациям, внедряющим BPM, от каких ошибок предостеречь? В частности, как вы обеспечили поддержку руководства и персонала, как справлялись с сопротивлением в ходе проекта?*

Основные сложности проекта заключались в отсутствии экспертизы внутри компании. Мы поручили разработку решения подрядчикам, но многие вопросы по отладке сервиса уточнялись уже в процессе работы.

## **Рекомендации**

- Выбирайте подрядчика с опытом в решении подобных задач.
- Привязывайте стоимость работ к полученному в итоге проценту распознавания. Это будет стимулом для подрядчика и позволит вам не переплачивать за решение, которое не даёт должной отдачи.
- Закладывайте ресурсы на создание своей системы проверки качества работы сервиса, а также на её отладку и развитие. При опоре только на сведения от подрядчика вы рискуете получить сырое решение, принцип работы которого будет непонятен для вашей собственной команды — следовательно, добиться хороших результатов будет сложно.
- Конкретизируйте способ оценки результативности в договоре — по ходу работы мы пришли от оценки качества распознавания заявки целиком к оценке каждого из полей заявки в отдельности. Качество распознавания различных полей отличается — для отладки необходимо сразу выделять те поля, что распознаются хуже всего.
- Проверяйте решение на новых данных. Те данные, на которых обучали модель, могут дать хороший результат. Но работать модели придётся с новыми данными —

подлинная оценка качества распознавания возможна только на том, что ещё «не знакомо» модели.

- Учитывайте влияние человека на данные при их обработке. Например, наш менеджер при заполнении формы заказа мог заменить написанное в письме «пл.» на «площадь» — а при проверке качества распознавания это воспринималась как ошибка. В процессе мы нашли более 200 подобных случаев при заполнении адресов. В итоге, проверку решено было сделать не по строкам адреса, а через определение координат обоих вариантов адреса с помощью геокодинга Гугла — при совпадении точек на карте распознавание считается корректным.
- Учитывайте влияние форматирования на качество распознавания. В какой-то момент мы обратили внимание, что форматирование письма может сказаться на результатах распознавания — присутствовала зависимость от отступов и пробелов. Потребовалось ввести специальную обработку, чтобы нивелировать этот эффект.
- Даже при частичном распознавании данных необходимо принимать их на проверку — некорректно целиком блокировать выдачу таких результатов. Для отладки важно видеть картину в деталях — даже малая часть распознанных данных сообщает о положительном результате, в отличие от сообщения о том, что заявка не распознана совсем.

*Результаты — в чем ваше главное достижение? Как проект отразился на выручке, затратах, качестве, продолжительности бизнес-процесса, удовлетворенности клиентов, управляемости, культуре организации и т. п.? По возможности, приведите цифры по состоянию на начало и конец проекта.*

- После внедрения сервиса среднее время обработки заказа сократилось с 5 до 4 минут.
- Это позволило обработать увеличение заказов на 10% в течение месяца без найма дополнительных сотрудников.
- Затраты на отдел по обработке заказов сохранились на том же уровне, что был до запуска сервиса.
- Кроме того, внимание сотрудников теперь сконцентрировано на проверке данных, а не на их копировании — ошибки на этом этапе сократились на 30%.

*Информационные технологии — какое ПО вы использовали (разработали, внедрили)? Как его интегрировали в существующую ИТ-инфраструктуру*

Пока продолжается отладка, сервис работает на стороне подрядчика. С помощью API он связан с внутренней системой компании.

*Партнер — кто из внешних консультантов, интеграторов или вендоров внес самый значительный вклад в успех проекта? Кратко опишите его роль.*

Сервис целиком разработан компанией «Экспасофт».