

## CASE STUDY

### ABBY FINE READER HANDPRINT

## FINE READER РУКОПИСЬ РАБОТАЕТ НА ФЕДЕРАЛЬНУЮ НАЛОГОВУЮ СЛУЖБУ

Можете ли Вы себе представить страну, в которой понятия не имеют, что есть налоговая декларация, а на налогового инспектора смотрят, как на заморскую диковинку? Именно такой страной была Россия вплоть до 1996 года, когда появился «Закон о налогообложении физических лиц», по которому граждане должны были предоставлять в налоговую службу декларации и справки о доходах. Закон должен был вступить в силу в 1997 году. К 1998 году объем подаваемых деклараций и справок о доходах составил 4 000 000 экземпляров. Учтывая что одна декларация содержала в среднем 7 листов, общий объем составил 28 000 000 страниц или 30 полностью забитых пятитонных трейлеров (только по Москве и Московской области). Все поступившие данные надо было ввести в базу данных Федеральной Налоговой Службы (ныне Министерство по Налогам и Сборам). Ввести в компьютер такое количество информации вручную к сроку было практически нереально. Ведь тогда понадобилось бы нанять тысячи сотрудников, которые в течение многих месяцев занимались бы черновой работой по «набивке» деклараций вместо того, чтобы заниматься сбором налогов с платежеспособных граждан, тем самым обеспечивая зарплату врачам, учителям, ученым и т.д.

К этому времени компания АБВУ уже реализовала проекты по автоматизации ввода в компьютер больших массивов информации в нескольких региональных отделениях Федеральной Налоговой Службы. И в 1997 году представители Федеральной Налоговой Службы обратились к АБВУ с предложением заняться проблемой ввода налоговых деклараций.

Надо отметить, что в России внедрение подобного рода проектов сталкивается с большими трудностями как технического, так и организационного характера. При всем том, благодаря (серьезным? Слово «героический» какое-то пафосное) усилиям департамента информационных ресурсов Налоговой Службы России совместно с разработчиками компании АБВУ, первая система автоматического ввода данных была разработана и внедрена менее, чем за полгода.

### Описание проекта

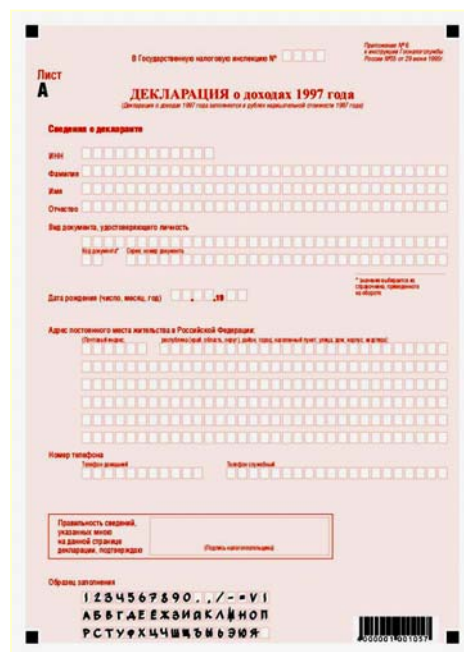
Перед АБВУ встала задача не просто организовать ввод информации в базы данных. Сложность была в том, что никогда ничего подобного в России (да и в мире) не делалось. А посему необходимо было продумать и организовать все процедуры, подготовить все необходимые материалы и инструкции от начала до конца, начиная с того, как будут выглядеть налоговые декларации и каким образом они будут попадать к людям, которым их предстоит заполнять, и заканчивая тем, как они будут архивироваться. Специалистам АБВУ предстояло разработать, помимо самого комплекса ввода данных, множество мелочей, играющих, тем не менее, не последнюю роль при реализации



проекта. Так, например, дабы избежать порчи документов во время перевозки из региональных налоговых инспекций в центры обработки информации, были разработаны специальные папки и сумки. Надо было продумать удобство и понятность декларации для заполняющих, что делать с оставшимися незаполненными бланками. Проект носил поистине революционный характер, ибо ничего похожего раньше не было не только в России, но и во всем мире. Задача стояла слишком серьезная, чтобы можно было допустить хоть одну ошибку. Для того, чтобы стало понятно, насколько разработанный ABBYU проект удачен и пригоден для внедрения на федеральном уровне, было решено сделать его «обкатку» в Москве и Московской области.

## Пилотный проект.

Для начала следовало разработать сам бланк налоговой декларации. Для автоматического ввода данных требовалась машиночитаемая форма декларации. Единственной организацией, в которой можно было увидеть машиночитаемые формы, на тот момент был Пенсионный Фонд, где они использовались для сбора сведений о доходах граждан для начисления пенсий. Налоговая декларация стала первой многостраничной машиночитаемой формой. Любая форма документа должна быть утверждена во множестве инстанций, прежде чем она получит официальный статус. Причем, в случае с обыкновенными формами, утверждается не только содержание, но и то, как бланк будет выглядеть. В нашем случае поскольку внешний вид бланка влиял на его машиночитаемость, решили утвердить только содержание бланка, а расположение полей отдать на откуп региональным отделениям налоговой службы. Содержание машиночитаемой формы было утверждено, и компания ABBYU приступила к разработке макета декларации для Пилотного проекта в Москве..



Во-первых, в качестве фона формы был взят красный цвет. Все надписи к полям и реквизиты формы тоже были выполнены в красном цвете – чуть более темного оттенка. При сканировании на сканере с красной лампой фон и все надписи на форме становятся прозрачными, и на изображении формы остается только написанный черной или темно-синей ручкой текст, а также черные квадраты и штрих-код – реперы формы, благодаря которым на страницу декларации при распознавании автоматически накладывается нужный шаблон и компенсируются возможные перекосы и линейные искажения изображения.

Во-вторых, поля формы состоят из отдельных маленьких прямоугольников, по одному для каждого символа. Это сделано для того, чтобы исключить слитное написание текста в полях.

В-третьих, каждая страница обладает уникальным набором реперов: помимо черных квадратов, расположение которых одинаково для всех страниц, каждая страница декларации имеет уникальный штрих-код и пронумерована уникальной буквой. Эти реперы позволяют не только устранять искажения и сдвиги изображения при сканировании, но и автоматически выбирать и накладывать

нужный шаблон на отсканированную форму, что делает возможным распознавание многостраничных форм.

Итак, макеты машиночитаемых форм деклараций были готовы к использованию. И тут возникли некоторые трудности.

Выше уже упоминалось, что формы деклараций были многостраничными. Поэтому возникала проблема, как при вводе в базу данных связать между собой страницы декларации одного и того же человека. Выдвигалось несколько интересных предложений, в том числе и идея присвоения ИНН (индивидуальный номер налогоплательщика), который бы служил идентификатором каждого заполняющего лица, но за отсутствием времени на проработку и внедрение всех этих идей, оставался только один выход – перевозить декларации со всей возможной аккуратностью, чтобы во время транспортировки, страницы не перемешались. Чтобы облегчить перевозку форм, и были придуманы специальные папки и сумки.

Кроме того, еще одну сложность представляли собой люди, впервые столкнувшиеся с необходимостью заполнять машиночитаемые формы. Часто они делали ошибки, поэтому в обязанности служащих налоговых инспекций входила проверка правильности заполнения декларации.

## Технология централизованной обработки сведений о доходах.

Эта технология была выработана Федеральной Налоговой Службой с помощью специалистов компании ABBYY.

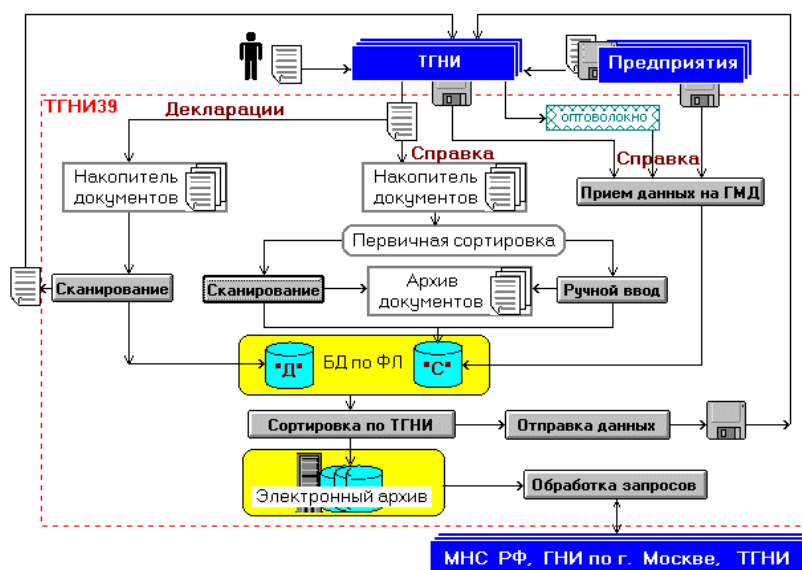


Рис. 1. Функциональная схема обработки документов в территориальной налоговой инспекции

Сведения о доходах граждан хранятся в единой базе, своей для каждого региона. В базу данные поступают по следующим каналам:

- ручной ввод данных с бумажных документов
- автоматический ввод данных на комплексах сканирования
- ввод данных с электронных носителей (сопровождается бумажной формой)

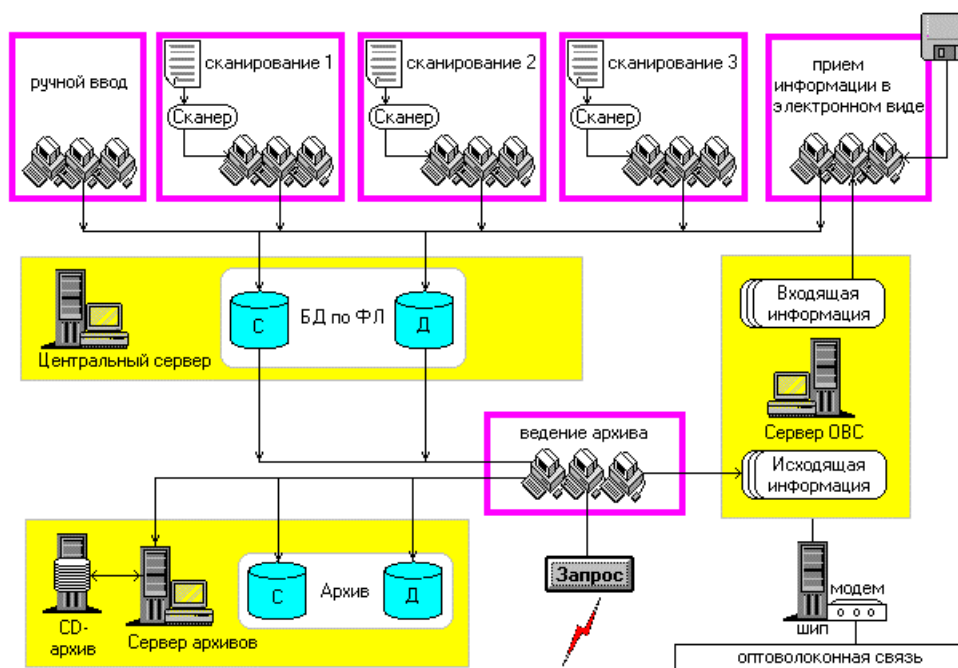


Рис. 2. Технологическая схема.

Периодически данные архивируются и хранятся в специальных архивах, откуда они при необходимости могут быть извлечены.

## Первые десять тысяч страниц.

С января 1998 года первые декларации начали поступать в налоговые инспекции. К этому времени в налоговой инспекции № 39 уже был установлен комплекс "FineReader Налог" по вводу деклараций и справок о доходах на 33 рабочие станции. Этот комплекс позволял обрабатывать до 10000 документов ежедневно.

По закону граждане должны были отчитаться о своих доходах до конца апреля, и уже ближе к весне тоненькие ручейки бумаги превратились в огромный поток документов. В налоговую инспекцию №39 свозились документы со всей Москвы и области, и пилотный комплекс успешно справлялся с поставленной задачей.

Все поступившие декларации должны были быть обработаны в течение 120 дней после окончания срока подачи. Конечно же в регионах, где использовался ручной ввод данных это требование было невыполнимо. *"Преимущества системы, разработанной совместно с компанией ABBYY особенно важны, учитывая сезонность нашей работы"*, говорит Игорь Попов, бывший руководитель ТГНИ №39, а ныне начальник Московского центра обработки данных при Российском Министерстве по Налогам и Сборам, *"если бы мы решили ввести в базу данных 4 000 000 деклараций за 120 дней вручную, нам бы понадобилось более 1000 профессиональных операторов и столько же компьютеров, и более 6000 кв.м. площади. И что бы мы делали со всеми этими людьми и площадями в течение остальных девяти месяцев?"*



Игорь Попов,  
начальник  
Московского центра  
обработки данных .

## Результаты.

Первый успех повлек за собой стремительное распространение технологии по всей России. На 1999 год в бюджет уже 5 крупнейших регионов России были заложены средства на автоматизацию процесса обработки деклараций и справок о доходах с помощью разработанного компанией АВВУ «Комплекса “FineReader Налог”», а ТГНИ № 39 превратилась в современный мощный Центр обработки данных, оборудованный на сегодняшний день уже тремя комплексами по вводу данных, каждый состоит из более чем 30 рабочих станций. В этом Центре обрабатываются документы со всей Москвы и Московской области.

Скорость сканирования одного комплекса составляет в среднем 80-100 страниц в минуту (используются сканеры ВапсТес S-185S). Согласно официальным результатам тестирований, проведенных в 1998г один оператор обрабатывал в среднем 3-6 страниц в минуту, при этом, скорость верификации обычно росла по мере роста навыков оператора. Специалисты компании АВВУ постоянно развивают и совершенствуют систему, повышая качество распознавания и удобство, а, значит и скорость, работы операторов.

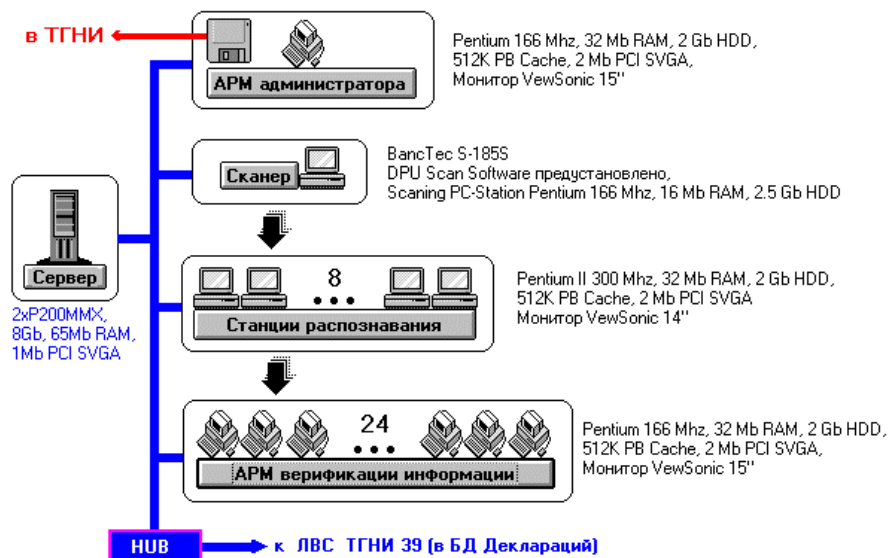
Эффективность и экономическая целесообразность системы автоматического ввода данных теперь не вызывает сомнений. Качество ввода данных при использовании “FineReader Налог” гораздо выше, чем при использовании ручного труда машинисток.

Без верификации FineReader Handprint допускает менее 5 ошибок на 1000 рукописных символов и менее 1 ошибки на 1000 печатных\*. Это в 4 раза меньше ошибок, чем профессиональная машинистка делает в начале рабочего дня. И в 6 раз меньше, чем та же самая машинистка делает вечером, в конце рабочего дня. Кроме того, программа автоматически проверяет входную информацию, используя справочники, базы данных, сравнивая сумму цифрами и сумму прописью и т.д.

К тому же автоматический ввод данных гораздо быстрее ручного. Этап верификации – единственный, на котором скорость работы ограничена возможностями человека. Поэтому, разработчики были особенно внимательны к этому этапу. Было предусмотрено несколько приемов и технологических возможностей, которые могут быть использованы как по отдельности, так и в комплексе, для организации эффективной верификации.

---

\* в аккуратно заполненных формах (при распознавании рукопечатного текста) и на текстах хорошего качества печати (для печатных документов)



**Рис. 3. Схема пилотного комплекса, установленного в Налоговой инспекции №39 г.Москвы в 1998.**

Трехступенчатая верификация FineReader, выделение ошибок в тексте, возможности автоматической проверки данных, позволяют сделать систему очень гибкой.

Благодаря всем этим технологическим возможностям, оператору необходимо лишь проверить неуверенно распознанные символы, помеченные системой. В результате, в день один оператор может обработать около 800 – 1000 страниц, тогда как при ручном вводе за день реально обработать лишь 250 – 300.

На сегодняшний день уже более 8 крупнейших регионов России используют автоматическую систему обработки Деклараций и Справок о доходах. Высокая скорость и отличное качество распознавания, автоматический контроль результатов и небольшая стоимость делают "FineReader Налог" наиболее эффективной системой автоматического ввода информации.

Одним из наиболее важных результатов внедрения автоматизированной системы обработки документов стало **повышение собираемости налогов в России**. Благодаря применению "FineReader Налог" Министерство Налогов и Сборов России стало одним из первых налоговых ведомств в мире, осуществивших автоматический массовый ввод налоговых деклараций.