

CASE STUDY

ABBYY FORMREADER

THE FORMS INPUT PROJECT IN THE STATE TAX SERVICE OF RUSSIA

Can you imagine a country with citizens who don't know what is a tax return and have never seen a tax officer? The Russian Federation before 1996 was an example of such a country. At that time Russia was making its first steps on the way to market economics, but a civilized country with market economics cannot exist without an effective tax service. So the law about natural persons taxation was passed by the parliament in 1996. According to this law, every citizen must submit tax returns and income reports (the so-called W-2 forms) to the State Tax Service, and the law was to come into force as of January the 1st, 1997. So in 1998 the overall volume of tax return and income report forms amounted to 4 000 000 documents from Moscow region citizens only. As there are 7 pages on average per tax return form, the amount of paper is 28 000 000 pages (30 five-ton full-loaded trucks are required to transport this pile of paper!). The data from these forms should be input into information databases of the State Tax Service (now the Ministry of Taxes and Duties). It went without saying that it was impossible to input such an amount of data in time manually, because it would require hundreds of operators doing an unskilled labor retyping forms for many months, as it was before with balance accounts input. And so they would have been typing in the forms instead of raising taxes from solvent citizens and thus guaranteeing the salary for physicians, teachers, scientists, etc., but an automatic solution was already on its way.

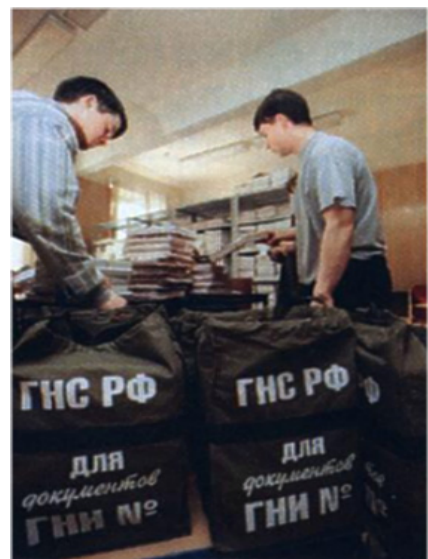
By that time, ABBYY had already developed a number of data input projects for local offices of the State Tax Service. So in 1997 the State Tax Service contracted ABBYY to find a solution for this data input task.

When one develops such projects in Russia, one usually encounters many difficulties of both technical and administrative nature that make the development age-long. But due to the extraordinary efforts made by Information Resources Department of the State Tax Service of Russia and ABBYY Software House the first automatic form data input system was developed and installed in less than 6 months.

The project details

The primary concerns of the developers were, naturally, the data input issues, but the scale of the project was much larger, because nothing of the kind has been done before in the State Tax Service. There were many organizational issues that should have been taken into account, from the question of where and how to print the forms through the procedures of their distribution among the local tax inspectorates to the questions of how and where to keep the forms archives and so on.

So the task of ABBYY was not confined to the development of the data input system; ABBYY was to develop the whole technology of data collection and input. And ABBYY performed most successfully. A single example: a set of special folders and bags to be used for transportation of tax return forms from the local tax offices to the data input centers was designed by ABBYY, in order to prevent the possible damage to the tax return forms. A lot of other technology



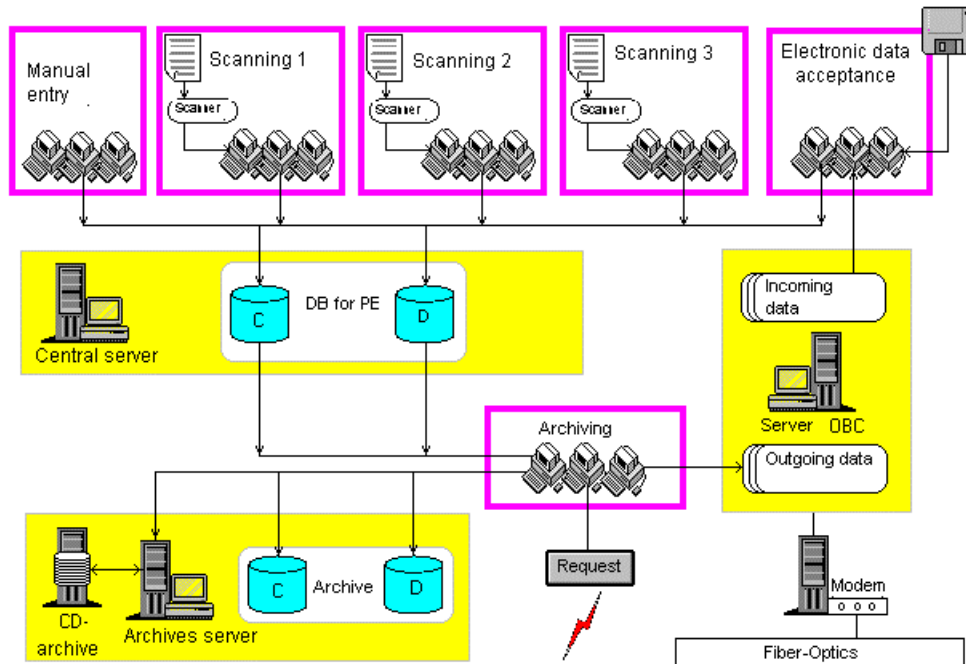
Technology of centralized tax return form processing

This technology was developed with the help of ABBYY experts.

All information about the income of each person is kept in a united regional database. The following channels are used to put information into the database:

- Manual data input from paper documents
- Automatic data input from scanning stations
- Data input from electronic media

Information is archived periodically and kept in special archives from which it can be easily got at any moment.



First ten thousand pages

In January 1998 the first tax return forms were collected by in the tax services. The FormReader system was already installed on 33 workstations in the Moscow Tax Office №39 by that time. The system could process more than 10000 forms per day. According to the law, all citizens were to report about their income until the end of April, so a huge river of documents began flooding the tax offices in spring. And they had only 120 days after May 1 to process all of them! Such time limits made it impossible to input all forms manually. But the FormReader system eliminated this problem for at least one Tax Office in Moscow.



Igor Popov,
Chief of the Data
Processing Center of the
Tax Ministry of Russia.

“The advantages of the system developed together with ABBYY are evident considering the season-dependent conditions of our work”, says Igor Popov, former Chief of Tax Office №39, now Chief of the Moscow Data Processing Center of the Ministry of Taxes and Duties of Russia, “If we tried to input 4 000 000 documents in 120 days manually, we would have to employ more than 1000 professional typists and to buy the same amount of computers, and rent the workspace of about 6000 m². And what we would have to do with all these people and workspace during the rest 9 months of the year?”

Results

The pilot project was declared a success, and the technology of automatic data input spread all over Russia. It was planned to install automatic form processing systems in 5 biggest regions of Russia in 1999. The Moscow Tax Office №39 became the Moscow Data Processing Center of the Ministry of Taxes and Duties of Russia. As of now, there are 3 systems of Automatic Data Input installed there, each consists of more than 30 workstations, and the tax return forms from almost every Moscow tax office are brought there for processing.

The scanning capacity of the system is 80-100 pages per minute on average. The scanner used is the BancTec S-185S scanner. One operator processes 3-6 pages per minute on average and the verification speed grows as the operator becomes more and more experienced. Further development and modification of the system performed by ABBYY’s specialists each year increase the recognition quality and make the verification process more fast and convenient.

Effectiveness and cost efficiency of automatic data input technology are now evident to everybody. The input quality featured by FormReader is much greater than the quality of manual input.

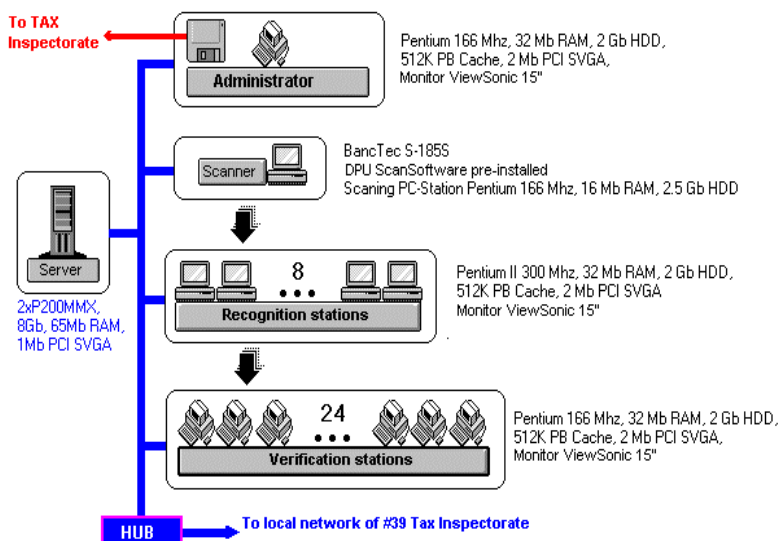


Figure 1. Pilot project scheme installed in the Moscow Tax Office №39 in 1998

Even without verification FormReader makes about 5 mistakes per 1000 handprinted characters^{*} and only 1 mistake per 1000 printed^{**} characters. It is 4 times less mistakes than a professional typist does in the morning and almost 6 times less mistakes than the same typist does in the evening. Besides, the application automatically checks input information, using reference files, databases, compares sum in figures and sum in words, etc., thus achieving the 100% reliable recognition results. Besides, it goes without saying that FormReader is faster than any, even super professional, typist. Verification is the only stage when the throughput of the whole system is affected by human productivity. That is why ABBYY's experts paid special attention to this stage. We provided a whole set of tools and techniques to be used separately or in combination to organize efficient verification. By changing template settings and adding rules it is possible to fine-tune the verification process to minimize human efforts and solve specific tasks. The three-step verification technology implemented in FormReader (group, context, and in-form verification), customizable error display level and checking level, etc. provide the flexibility that will help you to build the most efficient processing technology.

Thanks to all these tools, an operator needs only to check some uncertainly recognized characters. It allows to process 800-1000 pages per day instead of 250-300 pages per day in case of manual input. As of now, 8 largest regions of Russia make effective use of the enterprise version of FormReader system. High form processing speed, perfect quality of recognition, automatic control of recognition results and low cost make FormReader the most efficient automatic data capture system.

One of the most important results of this project is the increase of the total sum of taxes raised. The FormReader made the State Tax Service of Russia one of the first tax services in the whole world to use an automatic bulk input system of tax return forms.

^{*} for neatly written characters without corrections
^{**} for documents with good printing quality