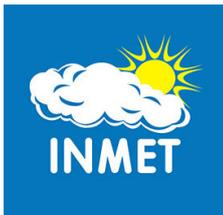




ABBYY FlexiCapture® Engine Helps to Read a 100-year Load of Climate and Weather Data

Customer Overview



Name

National Institute of Meteorology

Location

Brasília, Brazil

Industry

Government

Web

inmet.gov.br

Challenge

Extract data from scanned documents, dispersed in various patterns and forms, for a digital database containing historical climate data

Solution

ABBYY FlexiCapture

Results

- 3 million pages processed
- 4 billion characters recognized
- Increased speed and quality of recognized data

The National Institute of Meteorology (INMET) is an agency responsible for monitoring and forecasting weather and climate conditions in Brazil. It was established back in 1909 and since then its numerous monitoring stations scattered across the country have been scrupulously recording meteorological and climatic phenomena.

Challenge

Needless to say that INMET now possesses an impressive collection of weather records dating back to the end of XIX century. This data could be used to determine long-term weather patterns, analyze changing climate conditions and produce more accurate forecasts. It is hard to overestimate its importance for scientific and practical use, particularly in the activities of the Ministry of Agriculture, Livestock and Supply.

Unfortunately, until recently access to the meteorological database was complicated, if not impeded, by the fact that it existed only on paper. Obviously, it was really difficult to classify and search for particular pieces of information without the benefits of modern technology.

In 2012, INMET held a bidding process for the project to transfer information from paper (typed books in A3 format and handwritten booklets in various formats) to an electronic database in order to ensure continuity of the information and allow easy tabulation and analysis of meteorological data. The contract was awarded to Flexdoc Tecnologia da Informação Ltda, a reputed company specializing in providing end-to-end total solutions in the area of workflow process automation. The company develops software solutions for document processing, renders outsourced print services and provides integrated document management solutions, including document storage and BPO. Disposing of an area of 1500 m2 for document storage, it has implemented projects with volume exceeding 30 million treated documents.

The task set by INMET was a challenge due to its scale (more than 3 million pages had to be processed) and the variety of paper documents which differed in terms of structure, complexity and even physical condition. In the beginning, FLEXDOC didn't use any OCR tools relying solely on the templates pre-built in the system. However, after the first loads it became evident that significant variations in positioning of the fields hindered performance of the operators.

Solution

Recognizing the need to incorporate OCR in the process, FLEXDOC tested a number of solutions before settling on ABBYY FlexiCapture Engine. It was a perfect match combining exceptional data accuracy with advanced scalability. But the decisive feature was its flexibility as ABBYY FlexiCapture Engine provides a number of tools and utilities that allow easy 'zoning' of forms and ability to define and modify templates. This way all the range of documents produced over the decades could be processed in one system.

Firstly, scanned documents are imported and immediately sent to ABBYY FlexiCapture Engine for the recognition of form types and matching templates in order to localize each field. Automation of this step accelerates document processing a great deal as there are more than 20 types of brochures, each comprising of at least 6 types of pages. Some pages contain more than 150 fields.

The printed data are extracted using ABBYY OCR technologies. As for the fields with handwritten data, due to poor physical condition (some of the documents date back to early 1900s) and being machine-unreadable, it was sent to the operators who entered the data manually. By detecting the fields and their types, ABBYY FlexiCapture Engine greatly accelerated and facilitated the work of operators, who specialized in certain fields.

After that, all the fields go to verification. 100% accuracy is a must because the data will later be used for scientific calculations, researches and forecasts. In cases of discrepancies or doubts, the fields with unmatched results are sent to the supervisors for comprehensive analysis. Validation of conflicting situations and quality control before export still require participation of trained professionals specializing in meteorology.

Finally, metadata are exported into INMET's climate forecasting supercomputers. A total of 85 people are involved in the whole operational flow: from import, data extraction to verification and error handling.

"We needed a flexible solution because of the variety of scanned documents, dispersed in various patterns and shapes. The use of ABBYY FlexiCapture to recognize the fields' coordinates and exact location has brought us a huge performance gain when treating booklets."

Carlos Flávio Barreto F. de Souza, Director of Technology, Flexdoc Tecnologia

Results

Back at the beginning of the XX century INMET set an ambitious goal of recording and preserving all the information on weather and climate phenomena in order to use it for analyses and forecasts. However, as time went by this mission became virtually impossible as the ever-increasing amount of paper documents was making the data untraceable and inaccessible.

FLEXDOC was contracted to unlock these information treasures and transfer them to digital form. After testing other solutions they opted for ABBYY FlexiCapture Engine because of its successful references in the industry, scalability and flexibility, which allow processing of various types of documents.

Incorporation of ABBYY FlexiCapture Engine into the workflow has brought a significant increase in productivity. Thanks to ABBYY automation solution, FLEXDOC reduced the number of typists and specialists engaged in the project by 30%. Documents now are analyzed automatically, which makes the typists' tasks so easy that they could be assigned to a child. The job of qualified professionals is reduced to mere verification, which has enabled them to dedicate more time to creative and valuable activities. By the end of the project more than 3 million pages and 4 billion characters will have been processed. Thanks to ABBYY technologies, digitalization of all weather and climate information in Brazil, previously rendered as a long-term complicated endeavor, now will be completed by the year 2017.

About ABBYY

ABBYY is a leading global provider of technologies and solutions that help businesses effectively action information.

ABBYY 3A

Asia, Baltic, Middle East, South America, Africa
P.O. Box #32, Moscow, 127273, Russia
Phone: +7 (495) 783 3700
Fax: +7 (495) 783 2663
sales_3a@abbyy.com