

Nanjing Hydraulic Research Institute (NHRI) efficiently digitizes 60-year-long archive with ABBYY Recognition Server



“The primary requirement of NHRS for the hydrological testing documents processing is that the accuracy of data entry. Any input errors are intolerable. Scientific data cannot be any risk. ABBYY Recognition Server not only provides customers with the best accuracy of the input, but also greatly enhanced the efficiency of input, saving customers a lot of valuable time.”

Alex Chen, Director, DIT

Background

Nanjing Hydraulic Research Institute (NHRI), set up in 1935, originally known as the Central Hydraulic Engineering Research Institute, is the first of its kind in China. The institute is at present under the joint jurisdiction of the Ministry of Water Resources, the Ministry of Communications, and the State Power, P. R. China.

In 2008, NHRI received a mandate from the Ministry of Water Resources to input data from all the hydrological testing documents of the Chang Jiang River, dated from 1949 to 2007, into Nanjing hydraulic database. This data is used for future access and plays an active role in preventing the flooding of the Chang Jiang River.

The project

The foundation of the Nanjing hydraulic database basically went through two phases: Preparation phase and Implementation phase.

The Preparation phase consisted in collection, arrangement and analysis of relevant information. The Implementation phase was composed of using software to enter the information into the Nanjing hydraulic database.

The problem

After receiving the task NHRI started analyzing the data that needed to be input in the database. When the institute finally found out the amount of data to be processed, it was shocked. There were a total of more than 6,000 documents. Each document included several hundred pages collected during almost 60 years and representing all the hydraulic testing data. In accordance with the number of personnel and the speed of manual entry, input all of the data (including its proofreading) would have needed almost 5 years. It far exceeded the 2-year time limit set by the Ministry of Water Resources.



About Nanjing Hydraulic Research Institute (NHRI)

The institute is a multipurpose national hydraulic research complex, which, with the basic theoretical research and applied research as its major orientation, undertakes directional, principal and comprehensive researches for water conservancy, water resources management, hydroelectric power, waterway transportation projects, and the application and extension of small hydro power technology. For more information, please visit www.nhri.cn

To complete that task in a short period of time with high quality, the project team members began to explore existing OCR technology in Chinese market. They searched for a software solution that would provide high recognition accuracy, reliable digital identification, layout retention, strong technical support services and expansion capacity.

After careful testing some Chinese OCR software, NRHI was not satisfied with the digital identification. All documents had a chaos of crossing table lines and digits, which led to great difficulties for further proofreading and data compiling.

统计表 mm

日	一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
1						13.8	14.1	15.4	0.3			0.3*
2	0.8*					19.7	6.6	1.8				
3		0.4*				0.4	1.0	2.2				
4		0.3*				2.7	20.9	45.8				0.1*
5	1.3*			6.2	21.7		4.6	12.7			0.3	0.1*
6	0.4*		0.1*	12.1	3.3	0.6	7.7			1.5		
7					0.2	4.4	2.2			6.4		
8					0.3	0.3	15.8	20.5				
9				13.6	3.5	0.3	15.8	20.5				
10					4.0	6.2	12.5					
11					2.5	0.2	7.7					0.1*
12						2.5	5.5					
13						2.5	5.5			4.8		
14						19.8	0.2			0.8		
15		1.4*				0.8	0.2					
16	0.4*	0.6**		4.5	11.4	0.1		3.6	24.6	0.3		
17		0.3*			0.3			70.4				
18	2.1*				61.1							
19					26.6**							
20						4.4				6.4		0.2*
21		1.2*		2.5	12.0							
22		4.0*		0.2*		0.1						
23					7.3	4.0		1.0				0.3*
24					1.8	3.5						
25					2.4			1.0				
26					29.5	25.2		0.3		13.4	1.0**	
27	5.3*				0.4	0.4		0.9		0.3**	2.4**	0.2*
28												
29					4.0							3.2*
30												
31												
全年	7.1	9.8	1.9	26.8	135.0	165.8	171.7	224.7	27.9	35.1	6.9	1.3
统计表	4	7	5	9	12	22	20	14	6	7	4	7
统计表	3.3	4.0	0.7	25.6	81.1	29.5	48.9	70.4	24.6	13.4	3.2	0.3
统计表	统计表 (717.4)											
统计表	统计表 (74)											
统计表	1	1	3	7	15	30						
统计表	70.4	83.7	136.4	232.3	320.0							
统计表	8-17	8-4	8-4	7-27	7-23							

The solution

ABBYY partner in PR China, Digital Information Technology Co. Ltd. (DIT), found out about that situation and contacted NHRI to demonstrate ABBYY Recognition Server 2.0, a robust and powerful server-based OCR solution for automating high volume document conversion processes across enterprises. The proposed solution provides efficient tools for automatic and manual verification. For the latter, it includes client-side Verification Station designed for proofreading and correcting recognized texts. Recognition Server 2.0 can be configured to apply verification to all recognized pages, to selected document types, or to pages with questionable quality only.

NHRI tested the recognition accuracy through different resolutions and scan settings, in order to gain a better understanding of the product. ABBYY Recognition Server 2.0 appeared the only product that satisfied the NHRI team.

To improve the document conversion effectiveness and speed, NRHI decided to procure high-speed scanners and updated computer hardware. Then DIT helped NRHI to install and test the software and delivered its in-depth training.

The results

The superior recognition accuracy and layout retention of ABBYY Recognition Server 2.0 has won the customer acclamation. It has greatly cut down their labor costs and improved efficiency. By using ABBYY Recognition Server 2.0, NHRI is able to reduce the number of operators from 10 to 3, while increasing the throughput from 50 to 250 documents per day. As a result, ABBYY Recognition Server has been recommended to the Chinese Ministry of Water Resources.

ABBYY Russia

Russia, Baltic States, Asia, Middle East, South America, Africa, Australia & NZ
 P.O. Box #49, Moscow, 129301, Russia
 Tel: +7 495 7833700
 Fax: +7 495 7832663
 sales@abbyy.com

www.ABBYY.com

©2008 ABBYY Software House. All rights reserved. ©1987-2003 Adobe Systems Incorporated. Adobe PDF Library is licensed from Adobe Systems Incorporated. ABBYY, the ABBYY logo and FineReader are registered trademarks of ABBYY Software Ltd. Adobe, the Adobe logo, the Adobe PDF logo and Adobe PDF Library are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries. Microsoft, Excel, Windows are registered trademarks of Microsoft Corporation in the United States and/or other countries. Fonts Newton, Pragmatica, Courier ©2001 ParaType, Inc. Font OCR-v-GOST ©2003 ParaType, Inc. All other trademarks are the property of their respective owners. Information contained in this document is subject to change without notice.
 Part# 3505e



About Digital Information Technology Co. Ltd.

Digital information technology solutions provider, currently all-round cooperative partnership with ABBYY in China. Specializes in ABBYY channel distribution and giving end-to-end solutions in various fields like imaging, scanning, project consultancy and doing developments around ABBYY technology. DIT China has experienced technical skills in implementing ABBYY FineReader FlexiCapture and other ABBYY products. Focusing on different areas of challenges and needs concerned by customers. Committed to provide customers with up-to-date and competitive digital information technology solutions and services. Continuing to create the most value for the customer. For more information, please visit www.digitalitcd.cn

