

Shanghai International Studies University Reveals Its Multilingual Content to Academic Community Thanks to ABBYY Recognition Server 3.0



“ABBYY Multilingual OCR solution is the best OCR tech for Shanghai International Studies University (SISU), which provide best recognition performance and support largest quantity languages. Teaches sing highly praised on ABBYY product and improve their effeciney dramatically.”

Ren, IT manager, Shanghai International Studies University

Background

Shanghai International Studies University (SISU) is a research, teaching, and multidisciplinary university jointly administered by China’s Ministry of Education and Shanghai Municipal Government. Since its foundation in 1949, the University became one of the nation’s leading research centers. In 1996 SISU was granted by SEC (State Education Commission) to be a “Project 211” university. This is an initiative of the Chinese government aimed at strengthening of about 100 institutions of higher education and key disciplinary areas as a national priority.

Today SISU is the major center for foreign language materials and information consultation. It has over ten research institutions specializing in foreign languages and literature, international politics, economy, comparative cultures and cross-cultural communications. The University edits and publishes more than ten scholarly journals, including Foreign Languages, Foreign Language World, International Observation, Comparative Literature in China and Arabic World.

Three University Libraries boast a sizeable collection of more than 3 million volumes and nearly 4,600 newspapers and magazines in Chinese and other foreign languages. Apart from national works there is foreign literature in the original published abroad and reference books on language studies. Besides, University Libraries are subscribed to more than 1,200 newspapers and periodicals both in Chinese and foreign languages.

The core mission of the University is to create, preserve, and disseminate knowledge inside and beyond the academic community. For the past 60 years SISU has maintained a significant collection of research materials and resources which is constantly evolving to suit new and existing demands of the university work.

Challenge

To support research, teaching and scholarly communication SISU was actively applying new digital technologies intended to increase the availability of library collections and other materials. The central idea was to create a foundation for building a digital library and to provide the academic community with an easy access to primary source materials.

Looking for digital solutions in the fields of archival and preservation, SISU first decided to outsource its digitizing needs to a BPO company which had undertaken the complete document processing. It seemed to be a good idea until the document volume increased and a number of shortcomings were revealed: price was too high, productive capacity was scanty and texts recognition quality turned to be insufficient due to SISU versatile multilingual content.

Multilingual content always made a considerable part of educational material in SISU. Being a foreign language university it offers educational programs in different languages: English, Russian, German, French, Spanish, Arabic, Japanese, Persian, Korean, Thai, Portuguese, Greek, Italian, Swedish, Dutch, Indonesian, and teaching Chinese as a foreign language. Whereas very few BPO companies could really provide reliable support for recognition of all required



About Shanghai International Studies University (SISU)

Founded in 1949, Shanghai International Studies University, abbreviated as SISU, has developed into a comprehensive and multi-disciplinary foreign language university, influential both at home and abroad. Now it is one of the first key universities engaged in both foreign language teaching and cultural exchanges with foreign countries.

At present, the University is running 23 master’s degree programs, 10 doctoral programs and 1 postdoctoral station. Currently, SISU has a full-time enrolment of over 1400 postgraduate students, including more than 200 doctoral students.

For more information, please, visit www.language.shisu.edu.cn/

languages and effectively manage the growing scale of educational materials. On top of all this, a high outsourcing price didn't permit SISU to process the growing volume of books and magazines.

When SISU realized that outsourcing hadn't justified the expectations, the initial strategy was changed to the integration of book processing in the university home workflow system using a document capture technology. For this purpose SISU entered into a partnership with Digital Information Technology Co. Ltd (DIT), a software provider specializing in development of end-to-end solutions in the fields of imaging, scanning, and electronic document processing.

Project

The project primary goal was to make the academic material more accessible to scholars and deliver more efficient search and use of it. By that time SISU already had its own digital library workflow system which was a shared digital repository for storing university library digital content. With integration of a new digitizing technology it expected to gain a full access (for example, for reading and printing) to a vast body of materials from university library and other sources. DIT designed a solution based on a preservation and access model that would convert the primary source material into an electronic full-text searchable form.

The core of the solution was ABBYY Recognition Server 3.0, a robust and powerful server-based solution for automated OCR with scalable architecture. Featuring high-volume OCR and document conversion it ideally suited for document processing across the university large departments. With a single centralized OCR administrated just from one machine all recognition and conversion tasks were distributed among the processing stations and CPUs, balancing the workload across the system resources. The versatile recognition technologies of Recognition Server encompassed all essential parts of the document capture process in the university: scanning, recognition, document separation, classification, indexing, and delivery.

Built-in ABBYY OCR engine delivered the unprecedented recognition accuracy and ensured a reliable and accurate document processing. ABBYY Recognition Server 3.0 supports 198 languages with Latin, Cyrillic, Asian and other writing systems, like Chinese, Japanese, Korean, Hebrew, and Vietnamese. Such extensive language coverage permitted the university to process all required materials including multilingual documents. Besides, thanks to a prominent feature of ABBYY Recognition Server, an advanced recognition technology ADRT, which builds a logical model of the document, the structural parts and formatting elements were automatically identified and reproduced. Thus, users gained an ability to see the pages exactly as they appeared in the originals, complete with illustrations, charts and photos.

Thanks to the efforts of DIT and integrated ABBYY software SISU could eventually accomplish its task on a scope and scale that could not have been achieved previously by outsourcing.

Results

After SISU had integrated ABBYY Recognition Server 3.0, it not only could digitize a significant body of its academic material previously available as print editions. The full content (text and images) of books, journals, newspapers, and magazines had become accessible due to the acclaimed OCR technology provided by ABBYY software. Such approach to digitization had significantly eased the use of documents for scholars by adding advanced full-text search features which made all texts searchable and discoverable.

Thanks to unmatched scalability, ABBYY Recognition Server 3.0 was able to cope with any volume of documents, processing them as scheduled or round-the-clock. Vast language support enabled conversion of any item from the university collection regardless of the source language. SISU could eventually receive the desired characteristics:

- Cost reduction (in comparison with BPO services)
- High-volume processing
- High recognition quality and multilingual support

SISU was fully satisfied with the recognition results and now is planning to continue digitizing its documents with ABBYY Recognition Server 3.0. Moreover, ABBYY high performance stimulated the university to extend their project to its research institutes and towards additional support of teacher's work.

ABBYY 3A

Asia, Baltic, Middle East, South America, Africa
P.O. Box #32, Moscow, 127273, Russia
Tel: +7 495 7833700
Fax: +7 495 7832663
sales@abbyy.com

www.ABBYY.com



About Digital Information Technology Co. Ltd.

Digital Information Technology is a solutions provider which currently has an all-round cooperative partnership with ABBYY in China. The company specializes in ABBYY channel distribution and giving end-to-end solutions in various fields like imaging, scanning, project consultancy and doing developments around ABBYY technologies. DIT China has experienced technical skills in implementing ABBYY FineReader, ABBYY FlexiCapture and other ABBYY products, focusing on different areas of challenges and needs concerned by customers. DIT China is committed to provide the customers with up-to-date and competitive digital information technology solutions and services to create the most value for the client.

For more information, please visit www.digitalitcd.cn/

ABBYY®