

### DjVu Technology Inc. Developed Unmatched DLP Solution Using ABBYY FineReader Engine 9.0



*“High accuracy of OCR result is crucial for the development of information security system. Though 3rd-party OCR engine was bundled with our DjVu product, ABBYY FineReader showed superior accuracy. In addition ABBYY FineReader Engine processed documents in any languages, so ABBYY FineReader Engine was the perfect solution to integrate with our DjVu solution. We are considering expanding our business to fax information search solution using ABBYY FineReader Engine.”*

*Heungsik Choi, CTO of DjVu Technology Inc.*

#### The Background

Since information is the most valuable resource in organization, the risk of confidential data loss has become the critical issue for many companies. Rapid technological development of different communication channels (e.g. IM, USB, mobile phones) has caused tangible growth of unauthorized data loss outside the company. There are many ways for confidential data or proprietary secrets to leave the organization: e-mail, thumb drives, instant messages, webmail, new mobile technologies, HTTP and FTP links and many more.

The need for important corporate information protection has risen from the increase in theft and misuse of sensitive data as well as toughening of the compliance regulations. A recent trend toward transparency requires companies to allow information sharing with customers, business partners and suppliers. Therefore, the protection of confidential data from either malicious or accidental leaks has become one of the first-priority business security challenges which organizations are facing today. It should be noted that the instances of large scale data loss are the result of employers' carelessness and inadvertent mistakes rather than intentional thefts.

In order to secure the content distribution and management various technological tools for data leakage prevention (also known as anti-data leakage products) were invented. Offering a wide range of compelling benefits to organization they deliver a better understanding of what sensitive information is, how it should be used and what is the way to prevent the loss of it. Anti-data leakage products are usually deployed throughout an organization intended to identify and classify sensitive data, monitor for the unauthorized data disclosure and take appropriate actions to prevent any data leak.

Therefore, when Hyundai Construction, a major construction company in South Korea, decided to protect its sensitive data and introduce a data security system it invested in the outbound content management solution. Complete data security solution was provided by DjVu Technology Inc. which designed a data classification scheme and a storage architecture system to keep the enterprise safe.



#### About DjVu Technology Inc.

Established in 1996, DjVu Technology has been developing high-rate image compression and high-speed digital content publishing technology for more than ten years. It is specializing in scanned image and digital document management, photo image compression, publishing and security.

For more information, please, visit <http://www.djvutech.com>

## The Project

About 50 MFPs were installed at Hyundai Construction to provide a centralized document management and streamline business processes. The concentration of corporate information passing through multifunctional hardware allowed the comprehensive coverage of all confidential data across the network. The capability of tracking the activity of printer and MFP enabled such maintenance of the security information system that would deter employers from the leaking of important and valuable information.

A Korean software integration and distribution company – DjVu Technology Inc. which specializes in the area of digital image compression, scalable image viewing and secure content access and management was responsible for implementation of the solution of data leakage protection. DjVu Technology Inc. had designed for Hyundai Construction an end-to-end sophisticated solution intended for preserving the confidentiality of corporate digital data.

The project was aimed at the creation of secure content distribution and management required from DjVu Technology Inc. a powerful and intelligent document recognition and data capture system based on optical character recognition (OCR) technology. Finally a versatile software development kit, ABBYY FineReader Engine 9.0, was chosen to combine convenient image processing tools, document layout analysis, advanced conversion and compression with high-quality recognition results. Thanks to DIOTEK Co., Ltd., ABBYY's partner in Korea and a software development expert for embedded applications, the OCR technology was seamlessly integrated into the general software architecture.

To ensure the robust and efficient information security system DjVu Technology Inc. developed the following mechanisms:

- storage of the system architecture for scanned and copied documents
- sophisticated imaging technologies integrated with powerful OCR toolkit
- classified key word monitoring
- consolidated OCR-search on processed data

Thus the main goal of the project is avoiding the leaks of company's confidential information. Offering a wide range of compelling benefits to organization they deliver a better understanding of what sensitive information is, how it should be used and what is the way to prevent the loss of it. The project was based on the MFPs operation deployed at enterprise-level. All scanned and copied documents were recognized by OCR technology integrated into the multifunctional printers. Document recognition was provided by DIOTEK Co., Ltd and was based on the award-winning ABBYY FineReader Engine 9.0, a powerful software development kit for state-of-the-art recognition and conversion software technologies.

The OCR application had to satisfy a number of requirements determined by the peculiarity of the client, such as technical support for three languages: English, Korean and Japanese, and document recognition of any page orientation.

ABBYY FineReader Engine 9.0 ideally suited the project offering a comprehensive OCR technology for above-listed languages and their combination in multilingual documents. It also offered a range of image processing tools which improved the quality of the document image for further recognition and archiving such as image scaling and clipping, creating previews, image rotation, lines straightening, mirroring and inverting. Autodetection of page orientation (90, 180, and 270 degrees) was essential in case of Hyundai Construction because of a large input of images when the page direction is unknown and can be different. The system can automatically detect the orientation of each page and correct it if needed.

After the digitization and processing, information was stored in a centralized database where it could be readily searched. Accurate OCR results allowed applying indexing and monitoring (when a specific keyword is printed, scanned, faxed, copied, or sent, the system will set up a notification).

At the final stage all scanned images and digital documents were converted to DjVu® format. It ensured the highest possible quality of image with the smallest document file size, delivering reduced storage requirements and improved access without compromising image integrity.

## About DIOTEK Co. Ltd.

DIOTEK Co., Ltd. is a leading developer of mobile software solutions for mobile phones and embedded devices. It is engaged in developing of software solutions for handwriting recognition, mobile dictionary software, OCR solutions, and mobile photo editing software. The Company also provides mobile barcode software and digital ink solutions.

For more information, please, visit <http://abbyy.diotek.co.kr/>

## The Workflow

Success of the project was stipulated by the integration of different technologies: OCR, DjVu image compression and the achievement of the final solution - Enterprise Search Engine Solution.

The overall project was implemented over the period of three month and involved the following consequent stages:

- 1) **Storage at unified database.** Every scanned or copied document image was stored at a separate folder named by day and time.
- 2) **Identification.** The system identified new images, dropped them to central storage and then forwarded them to recognition servers for further processing.
- 3) **Text recognition.** Installed on two servers, ABBYY FineReader Engine 9.0 performed a full-text recognition, turning digital documents into searchable and reliable formats and creating document archives. It offered language support for English, Chinese and Korean character recognition, along with processing of multilingual documents.
- 4) **Storage of recognized results.** OCR results were stored as text files in the destination folders.
- 5) **Indexing.** Search Engine monitored OCR destination folders to see if new text files arrived and started indexing.
- 6) **DjVu compression.** Images stored for OCR were compressed and stored by DjVu imaging serves to maintain file archives.

## The Results

Introduced for the first time at Hyundai Construction the system of document leakage protection, based on OCR technology, significantly reduced the number of breaches of information security and enabled creation of the unified corporate storage system. Document recognition of the highest level provided by ABBYY FineReader Engine 9.0 extended the monitoring of information with the data embedded inside images – thus covering the complete range of printable documents.

The key benefit was a decrease of expenses caused by unwanted information leaks. New data security system deterred employees from leaking important information, curtailed printing of sensitive documents and protected confidential data from misuse.

Finally DjVu Technology Inc. developed an unparalleled data prevention solution which provided the client with a regulatory compliance and enhanced security system.

### 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](http://www.ABBYY.com)

**ABBYY®**