ABBYY Recognition Server 4 Novosti

New Document Processing Scenarios

- Processing Read-Only Archives and Document Libraries – The administrator can set an existing archive folder or SharePoint library as a source of input documents, and the original collection of documents stored there will not be moved, deleted or altered. If the same folder or SharePoint library is defined as the source of input documents and the location for converted documents, files produced by ABBYY Recognition Server (for example, searchable PDF files) can replace the originals, be stored with a new file name or as a new version of the document, according to the preferred settings.
- Processing Files by Mask – The administrator can now select types of files (for example only TIFF and JPEG files) or define a mask for names of files that need to be converted. Files that do not conform to the mask can be moved to a different folder, stay in place or be copied directly to the output location. This allows creating searchable archives that completely "mirror" the structure and content of the original archive. The whole variety of documents from the original archive (including images, Word, PowerPoint and Excel files) will be recreated in the output archive; however the static document images will be replaced with their copies in PDF or another searchable format.

Enhanced Integration with Microsoft SharePoint

Converting Images Within SharePoint – ABBYY Recognition Server's integration with SharePoint has been extended beyond simple "capture to SharePoint" capabilities. Now it can process PDF files and images in existing SharePoint libraries and convert them into searchable PDFs, enabling SharePoint to include these documents in the search index.

Enhanced Set of PDF Features

- Smart PDF Processing – To avoid unnecessary processing of already-searchable PDF files (i.e., PDF files containing a text layer), ABBYY Recognition Server can skip such files or intelligently decide which of them need to be reprocessed via OCR. In the latter case, ABBYY Recognition Server detects the text layer in each PDF file, checks its quality and decides whether it should export the original file unaltered or replace the text in it with the newly extracted text. This intelligent algorithm allows PDF files that are already searchable to stay unaltered and shortens the overall conversion time.
- Support of Latest PDF/A Standards – In addition to PDF/A-1a and PDF/A-1b standards for long-term archiving of digital documents, ABBYY Recognition Server supports creation of PDF/A-2a, PDF/A-2b and PDF/A-2u files.
- Text Injection in Image PDF Files – The new version of ABBYY Recognition Server supports "injection" of a text layer in an image PDF file while preserving the original image quality as well as all properties of the source PDF, such as bookmarks, annotations, metadata and attachments.

Architecture and Administration Enhancements

- Enhanced Robustness – With the addition of an SQLite database and enhanced fault tolerance mechanisms, ABBYY Recognition Server has become even more robust and able to consistently handle high workloads and multiple clients within the same server installation. This makes the software perfectly suitable for extensive document processing projects running in an unattended mode, requiring minimum human involvement.
- Enhanced Logging – To make server administration even easier, the log of completed tasks has been enhanced with additional information about each individual processed document. The log retention policy can now be defined based on a time period or file size, providing the system administrator with more flexibility.
- Easy User Management via Active Directory – Permissions for tasks such as administration, verification and indexing can now be assigned via Active Directory user groups. Once a user is added to (or removed from) an Active Directory group registered in ABBYY Recognition Server, his or her permissions are automatically updated.
- Support for SSL – When sending documents to Recognition Server by e-mail, it is now possible to use SSL protocol for stronger security and data protection.

Enhancements to Operator Stations

- Import of Value Lists in the Indexing Station – The list of possible values for index fields can be imported from an external file.
- Possibility to Select Documents for Verification or Indexing – The operator can manually pick a document from the verification or indexing queue in order to process it ahead of others.
- Saving Intermediate Verification Results – In the process of verifying documents with large number of pages (for example when converting books), the intermediate verification results can be saved before the verification job is fully completed.

Other OCR Technology Enhancements

- Improved Recognition of Arabic – According to ABBYY internal tests, the new version of ABBYY Recognition enables up to 20 percent faster recognition of Arabic documents and up to 10 percent better accuracy on office documents and books in Arabic language.
- Recognition of Text Printed Sideways – Text printed sideways can now be extracted from documents (for example, construction plans) along with the normal horizontal text.