ABBYY FlexiCapture 9.0
Document Processing Guide

Dear user!

This guide describes the actions performed during document processing in a set-up project.

If you still need more information, try the following:

- Use other help files. You can open them using the program menu or by clicking “F1” or Start > Programs > ABBYY FlexiCapture 9.0 Stations> Helps.

We hope you will enjoy using our product!

Contents

1. Introduction

1.1. The Purpose of Data Capture

A large variety of documents is used today: in business, production and services. Applications, questionnaires, invoices, drafts and other documents are essential for any company. Modern information technology makes paper documents insufficient, and most data is converted to electronic form for storage, analysis and processing purposes.

The most labour- and time consuming thing about electronic documents was data input. It could only be entered by hand, which was reasonable with a small amount of information. However, this doesn't work well with large document volumes. The speed of manual entry cannot be momentarily increased when the situation so demands, because manual entry groups are hard to manage, and the costs of changes may equal those of starting the process anew.

Thus, manual entry is not the optimal way. Its alternative, a simpler and more effective way, is an automatic data capture system like ABBYY FlexiCapture 9.0.

1.2. Data Capture Automation

ABBYY FlexiCapture 9.0 is data capture software for processing structured, semi-structured (FlexiLayout) and non-structured documents.

Automatic data capture consists of the following stages:

- A pack of pages is **scanned** using a document scanner first;

- The scanned pages are then automatically bundled into documents;
- The characters are automatically **recognized**;
- Uncertainly recognized characters are sent to the Operator for **checking** (**verification**),
- Confirmed data is finally **exported** to a file or database, and document images are saved to the specified folder. Images can be saved in a graphical format or as a searchable PDF file.

ABBYY FlexiCapture 9.0 is an efficacious data capture automation solution allowing easy control over work progress and quality.

1.3. Documents You Can Process in ABBYY FlexiCapture 9.0

ABBYY FlexiCapture 9.0 is a data capture application supporting different document types.

The following document types can be processed in ABBYY FlexiCapture 9.0.

- ***Structured documents***. Documents with dedicated data fields that remain constant in quantity, position and formatting throughout the document copies are called structured. These forms are often issued in printed form for filling by hand. In order to identify a structured form and capture the data, a layout must be created indicating field locations to the program. Layouts are created in ABBYY FlexiCapture 9.0 during project setup.
- ***Semi-structured documents***. Documents with data fields that differ in quantity, position and formatting from copy to copy are called semi-structured or flexible. Invoices are an example of this type, because they are often different in the number of items and formatting, for they are issued by different companies. All invoices include an account number and the amount of payment, but these are located in different parts of the document. ABBYY FlexiCapture 9.0 uses FlexiLayouts for identification of semi-structured documents and data capture. FlexiLayouts are created in ABBYY FlexiLayout Studio. For information on ABBYY FlexiLayout Studio, see its Help File. Processing of flexible or semi-structured documents is different from that of structured documents only at creation and layout matching stages.
- ***Non-structured documents***. If you need to process non-structured documents with information presented in free form, for example contracts, letters, orders, diagrams, ABBYY FlexiCapture 9.0 will also be a solution. Non-structured documents can be automatically identified as supplements to structured or flexible documents, or by using FlexiLayouts, and then exported to image and searchable PDF files. Index fields can be captured from non-structured documents automatically (using FlexiLayouts) or manually. A typical non-structured document processing scenario would be converting a paper archive to electronic form, with capture of a couple of index fields required for attribute-based search.

1.4. Data Control

Recognized data is controlled both by Verification Operators and automatically.

Operators check and, if necessary, correct the uncertainly recognized characters. This process is called verification and is carried out in two stages.

- The first stage is group verification – grouping of character images recognized with the same value, and displaying them on the Operator's screen for him to notice and correct the characters that stand out from the rest.

- The second stage is field verification – verification of specific field values.

The system automatically performs the following data control operations:

- Checking if pages were correctly assembled into documents.
- Checking if the data corresponds to the specified format.
- Checking if the rules are met. Rules are certain interdependencies among document fields specified at project setup.
- Batch integrity check.

If errors occur during the checking process, the corresponding document (batch) is sent to the Operator for checking.

1.5. ABBYY FlexiCapture 9.0 Verification Station

The following stations are used for document checking and correction in ABBYY FlexiCapture 9.0:

- Data Verification Station – for group and field data verification;
- Verification Station – for all document checks.

2. Document Processing

Documents are processed in several stages automatically or by an Operator. All stages form a strict order, in which the previous stages are required to be passed for the current stage to begin. Basically, processing is divided into the following stages: image import, recognition, verification, export.

A processing task queue is formed at each stage. During processing, documents go from the initial queue to the final one – export. Separate queues for checking and correcting specific errors can be created at the verification stage to increase processing speed and quality.

Data import, recognition and export are carried out automatically. Thus, the following queues can be accessed by Operators:

- Document Assembly Check;
- Data Verification – group and field verification of uncertainly recognized characters and format errors;
- Verification – rule and format error verification (in the document window);
- Batch Integrity Check;
- Export Confirmation;
- Exceptions.

The queues that are accessible to an Operator are defined by the Operator's role. The Data Verification Operator can only access the Data Verification queue. The Verification Operator can access all queues, except for Export Confirmation and Exceptions. The Senior Verification Operator can access all queues.

The list of queues accessible to Operators is configured during project setup. The Verification and Exceptions queues are required to be selected. If the Document Assembly Check and Data Verification

queues are not selected, the corresponding processes are carried out within the Verification queue. If batch integrity is checked, but the Batch Integrity Check queue wasn't selected, batches with errors appear in the Exceptions queue. If the Export Confirmation queue wasn't selected, export is carried out right after verification ends.

Documents or batches with errors that occurred during recognition or export are also sent to the Exceptions queue.

3. Tasks

Documents are sent to Operators for processing in tasks. Depending on the queue and project settings, the number of documents in the task can vary:

- If batches are to be processed as a whole, all documents in each batch are added to the task. During Document Assembly Check, Batch Integrity Check and Export Confirmation, the batches are processed as a whole. The same can be done in the Verification queue, if configured in project settings;
- If it isn't necessary to process the batch as a whole, each task will contain 10 documents. If the number of documents in the batch is not a multiple of 10, the odd documents will form a separate task. Thus, you will get a number of tasks with 10 documents each, and, possibly, one task with 1-9 documents. The described procedure applies to tasks in the Data Verification queue and (with the required settings) the Verification queue.

A task only consists of documents from one and the same batch.

The Operator's work begins with the choice of queue and task reception (Figure 1).

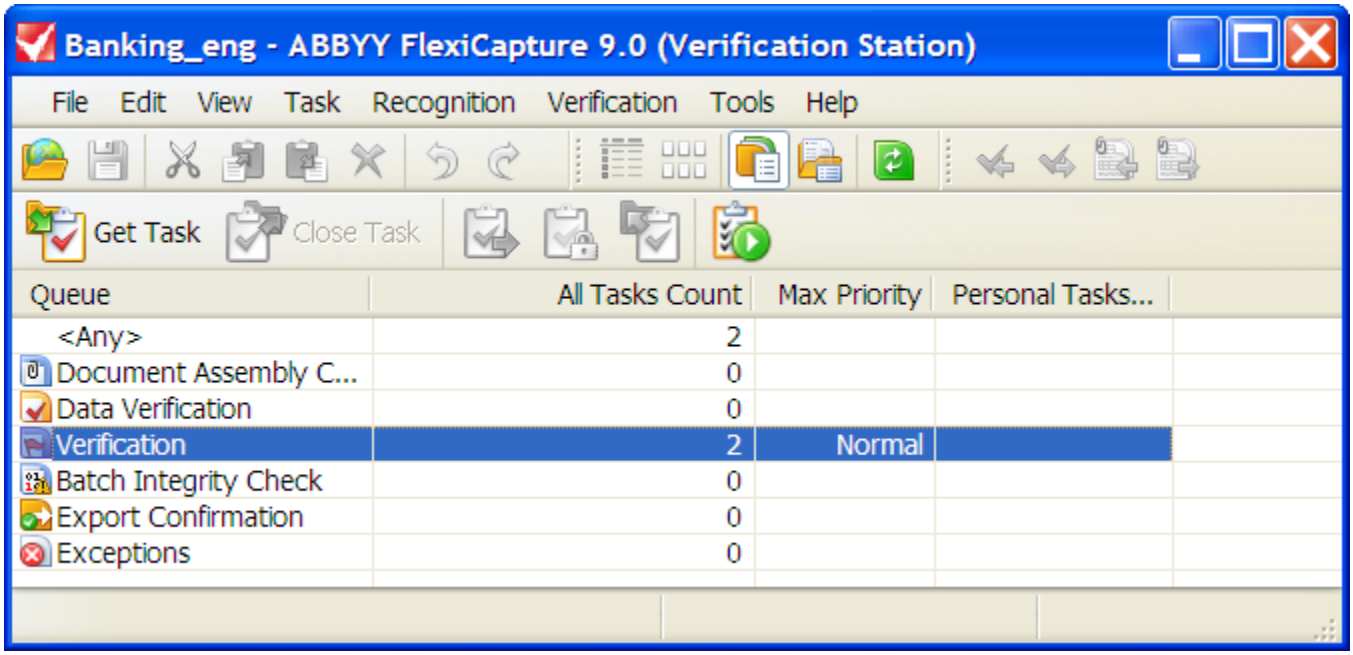
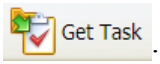
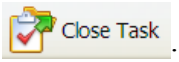


Figure 1. The main window

To get a task, click 

. When the task comes from the Data Verification queue, the system will open either the group verification or the field verification window. When the task comes from any other queue, the program will open the list of documents.

When task processing is complete, it must be closed by clicking 

. The documents (batches) will be then sent to the next processing stage. If the task was taken from the Data Verification queue, it will close automatically when data verification is finished.

The task can be cancelled with the  button (Reject Task).In this event, the task will be disassembled and returned to the queue it was taken from. You can also postpone a task by clicking ). The list of postponed tasks is displayed on the bottom of the main window (see Figure 1). You can open a postponed task by double-clicking it. Other Operators will be unable to check out a postponed task.

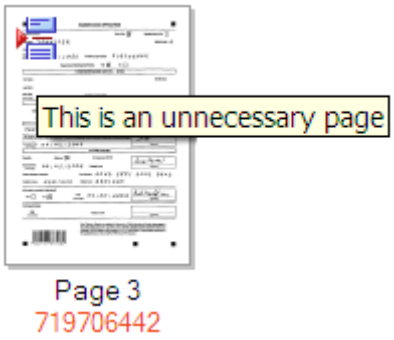
You can also send it to the Exceptions queue by clicking  if something unexpected happens during processing. The Senior Verification Operator can send the task to any queue as well as assign an Operator to it also by clicking this button.

The **Task** menu commands can be used to send some documents of the task to the Exceptions (**Send Documents to Exceptions…**), add documents to an existing Exception (**Add Documents to Exception…**), delete them (**Delete Documents from Exception…**), and view the lists of existing Exceptions (**View Exceptions List…**). In the Exceptions list, comments are shown for each Exception. Comments are added when adding tasks or documents to Exceptions. Tasks in the Exceptions queue are processed by the Senior Verification Operator.

4. Assembly Check

Document assembly is checked in multi-page documents. Pages are assembled into documents automatically. You only need to correct errors, if any occur.

If the page order doesn't match the one specified, or key field values on different pages don't match, the document will be marked red, and the corresponding error message will appear in the document window.



In this case, check that the pages weren't mixed up during scanning. Sometimes, assembly errors can be corrected by simply changing the page order.

Document assembly can be easily checked in page thumbnail view mode (Figure 2). You can alter page positions and even move them from one document to another using the drag-&-drop method. Thumbnail view mode also incorporates pop-up hints, as shown in the picture.

If verification by key fields must be carried out for correct document assembly, the key field values will be displayed under the image of each page (see Figure 2). If the key fields don't match within the document, field values will be marked red.

Key fields may mismatch because of faulty filling or recognition. Check the key field values. If the values still don't match, check if the pages belong to one and the same document. If page order is wrong, find the pages with matching key field values and assemble these pages into documents.

Note. You can zoom in and out of page thumbnails. To do it, hold Ctrl and roll the mousewheel up or down.

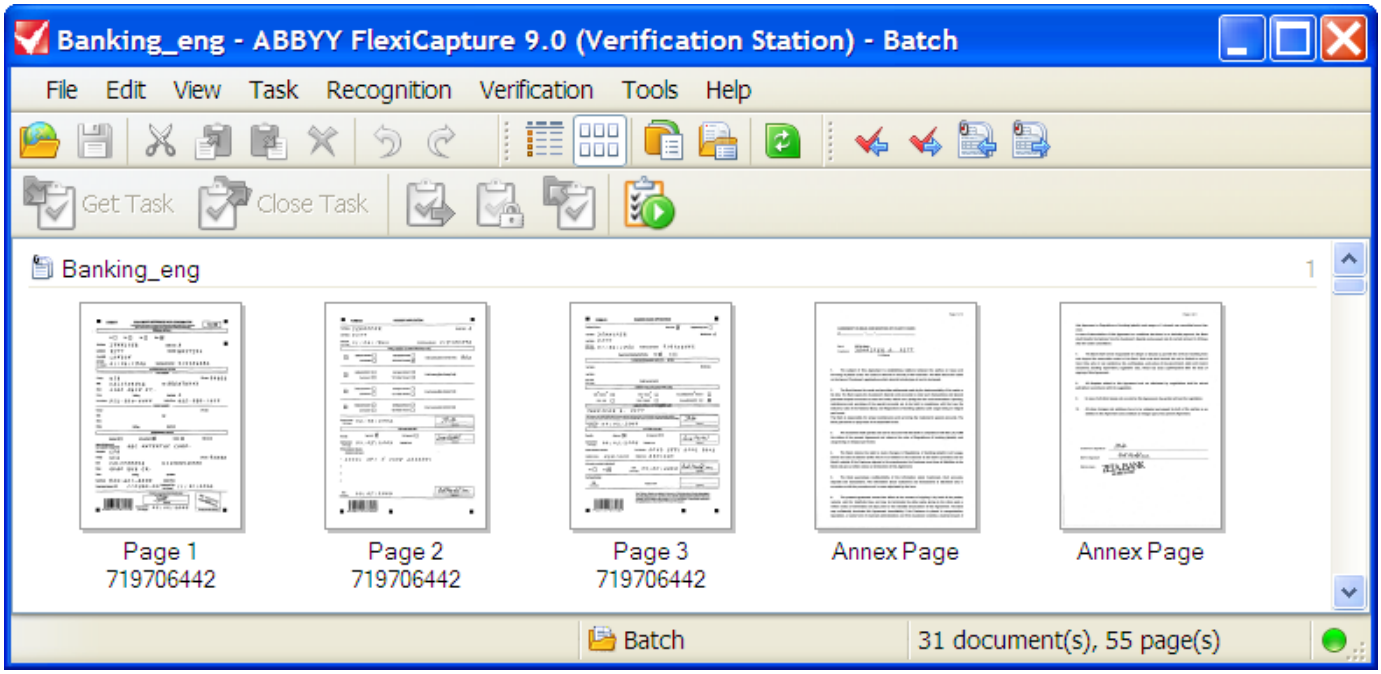


Figure 2. The main window in thumbnail view mode

Close the task when you are finished correcting assembly errors. If errors still remain, the system will issue a warning.

4.1. Adding Images

Images are imported from Hot Folders into the system automatically or by an Operator at the Scanning Station. In case of need, the Verification Operator or the Senior Verification Operator will be able to scan the missing images (**File > Scan Images…**) or load them (**File > Load Images…**). This may be useful for correcting assembly errors.

5. Data Verification

When a task is taken from the Data Verification queue, the Group Verification window will open. When Group Verification is finished, the Field Verification window will open. It will close automatically when the task is completed.

5.1. Group Verification

Group Verification is checking grouped character images that were recognized with the same value. All characters that were recognized identically (for example, the 0 digit on Figure 3) are displayed in the verifier to confirm correct variants and only leave incorrect or uncertain ones for the next stage.

If you are uncertain about the value of this or that character, do the following:

1. Select **Show Character Image** in the local menu of the selected character or press F2. The image of the field in which the character was found will appear.

2. Click **View > Field Image > Show Field Image** in the verifier window or press Ctrl+I. This will enable pop-up fields when moving the mouse pointer over the verified character.

If the character was recognized incorrectly:

1. Select the character that doesn't correspond to the group character and specify the correct value. The value will be marked green in the upper-left corner of the character image.
2. If you are uncertain about the character value even after checking the context, left-click the character to leave a red question mark. Character status can be also changed using the **Change Status** button.
3. Delete the character, if required, by selecting it and pressing **Del**.

Correctly recognized characters must be confirmed by clicking **Confirm**. To confirm all displayed characters, click **Confirm All** on the toolbar.

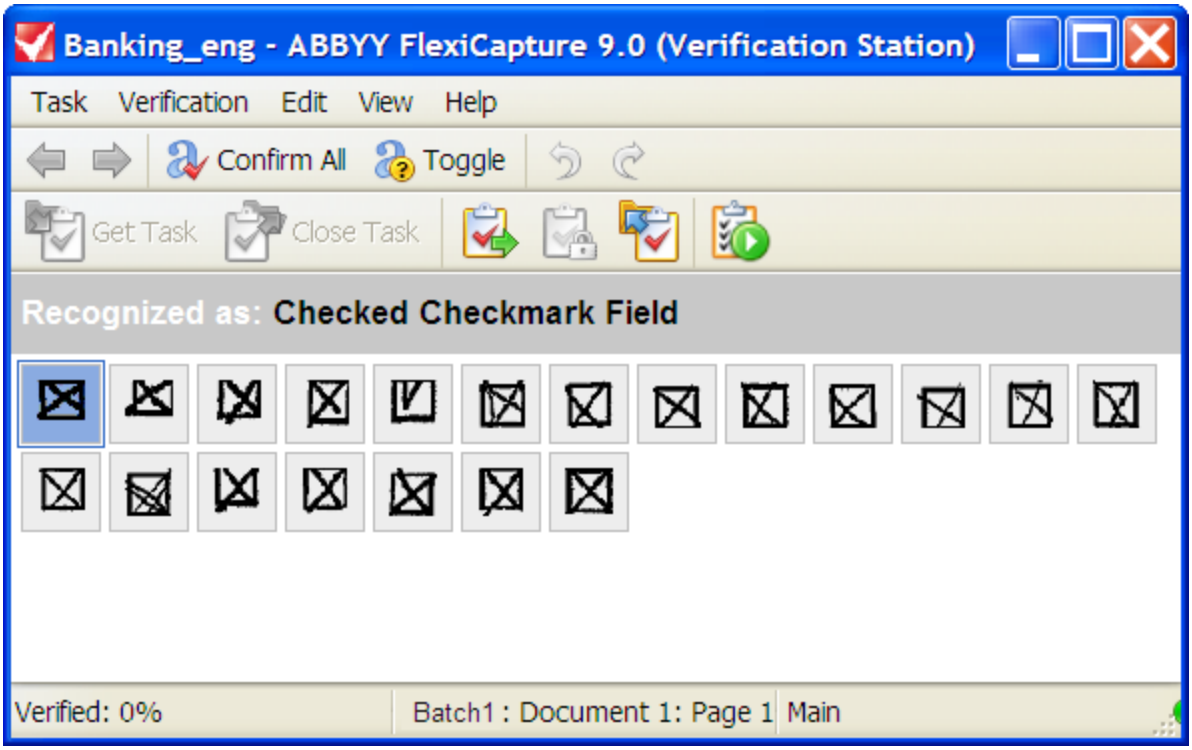


Figure 3. Group Verification of digits

5.2. Field Verification

Field verification allows checking uncertainly recognized characters in the context of a field with known value area. For example, we know what values the *Country* field can assume, so it will be easy to correct the field value, if the need to do so arises.

Standard word processor modes, such as character insertion and replacement, can be used to correct misrecognized symbols. You can switch modes by pressing **Insert**.

Check the recognition results for each field, one by one, make the necessary corrections and then confirm the value be pressing **Enter** or clicking **Confirm Field**.

Fields with values not corresponding to the specified data type are marked with a red **F**, and an error message appears. The same happens with the fields for which the rules were specified, but the field values

don't correspond to the rule. Correct the values of such fields. If you are unable to do the correction, postpone the value check by clicking **Postpone Field** to make corrections later at the Verification stage (in the document editor window).

Figure 4. Field Verification window

6. Verification

During verification, the following checks are carried out: the format check (recognized value doesn't meet the field format) and the rule check (if, during project setup, the rules were specified, which the field data must correspond to. If the Assembly Check queue and the Data Verification queue weren't separately specified, these processes are carried out within the Verification tasks (see the "Assembly Check" and "Data Verification" sections).

Format and rule errors are corrected in the document editor window. To open the window, double-click the name of the page or document. Recognized data or the document image may appear in the window. The document editor also incorporates an error window. You can customize the position of windows by clicking the **Layout** button.

Characters for verification are marked red, and the fields with incorrect data types and not meeting the rules are highlighted. To select the previous and the next verification objects, use the  and 

buttons. You will go through assembly errors, uncertainly recognized characters, rule errors, and other items one by one.

Rule issues are marked with flags: yellow for warnings and red for errors. Rule warnings and errors are displayed in a separate window of the document editor, and documents with rules that weren't met are marked with red flags in the document list.

If some of the rules haven't been met, the Operator must check the data verification results. If there are verification errors, correct them. If there are filling errors that cannot be corrected, send the document to the Exceptions queue.

After an error has been corrected, the rule is automatically checked once again. To repeat verification manually, click **Tools > Re-check Rules**.

7. Batch Integrity Check

During export, the Batch Integrity Check is carried out by a script that is specified by the Administrator. If there are errors, the batch is sent to the Batch Integrity Check queue (or, if this queue is absent, to the Exceptions queue). Task comments are specified by the Administrator in the batch integrity check script. In the resulting task, correct the errors that corrupt batch integrity and close the task.

8. Export and Export Confirmation

The Export Confirmation queue is available to Senior Verification Operators. When receiving a task from this queue, check if the documents of the task can be exported. When the task is closed, the Batch Integrity Check is carried out. If the check is successful, the documents are exported automatically. If the documents or the task are not to be exported, postpone the task or send it (or part of it) to another processing stage or to Exceptions.

If an error occurs during the integrity check, the system will issue the following warning: "Batch <batch name> contains integrity errors. Continue exporting?". Click **Yes** to continue, and **No** to cancel. If you choose to cancel the export, the batch will be sent to the Batch Integrity Check queue (or, if this queue is absent, to the Exceptions queue).

The Senior Verification Operator can also carry out the export locally. To do it, select the documents for export in batch view mode and click **File > Export to…**, and then choose the necessary command.

9. Working with Exceptions

Batches and documents with errors that occurred during recognition or export are sent to the Exceptions queue. If batch integrity checks are enabled, but no separate queue was created for these checks, batches with integrity errors will also be sent to the Exceptions queue. Verification Operators can also fill the Exceptions queue with tasks or documents with unforeseen errors that occurred during processing.

Tasks from the Exceptions folder are only accessible to Senior Verification Operators. When the errors have been corrected, the documents or the tasks must be sent to the required stage to continue the processing.

10. Keyboard Shortcuts

10.1. Main Window

Save	Ctrl+S
New Batch	Ctrl+N
Load Images	Ctrl+O
Scan Images	Ctrl+K
Import Images	Ctrl+I
Export Data to Files	Alt+Shift+S
Export to Database	Alt+Shift+D
Undo	Ctrl+Z
Redo	Ctrl+Y
Cut	Ctrl+X
Copy	Ctrl+C
Paste	Ctrl+V
Delete	Del
Select All	Ctrl+A
Find	Ctrl+F
Find Next	F3
Next Document	Ctrl+D
Previous Document	Ctrl+Shift+D
Despeckle Image	Ctrl+Alt+K
Invert Image	Ctrl+Alt+V
Rotate 90º Clockwise	Ctrl+W
Rotate 90º Counter-Clockwise	Ctrl+Shift+W
Batches	Ctrl+B
Queues	Ctrl+Q
Details View	Alt+1
Thumbnails View	Alt+2
Full Screen	F11
Refresh	F5
Properties	Alt+Enter
Document Editor: Image Mode	Ctrl+1
Document Editor: Data Mode	Ctrl+2

Zoom In	Ctrl +Num+
Zoom Out	Ctrl+Num–
Thumbnails Size: Increase	Ctrl+Shift+Num+
Thumbnails Size: Decrease	Ctrl+Shift+Num–
Get Task	Ctrl+G
Analyze	Ctrl+E
Match Document Definition	Alt+Shift+E
Recognize	Ctrl+R
Run Verification	F7
Next Item to Verify	F4
Previous Item to Verify	Shift+F4
Next Assembly Error	F9
Previous Assembly Error	Shift+F9
Next Uncertain Character	F8
Previous Uncertain Character	Shift+F8
Next Rule Error	F6
Previous Rule Error	Shift+F6
Update to Latest Version	Alt+Shift+U
Re-analyze	Ctrl+Alt+E
Re-recognize	Ctrl+Alt+R
Re-check Rules	Ctrl+F6
Check Batch Integrity	Ctrl+Shift+E
Help Topics	F1
Exit	Alt+F4

10.2. Group Verification Window

Confirm All	Enter
Postpone All	Ctrl+Enter
Change Status	Space
Next Page	Page Down
Previous Page	Page Up
Undo	Ctrl+Z
Redo	Ctrl+Y

Select All	Ctrl+A
Show Character Image	F2
Full Screen	F11
Show Field Image	Ctrl+I
Field Image: On Top	Alt+1
Field Image: On Bottom	Alt+2
Zoom In	Ctrl+Num+
Zoom Out	Ctrl+Num–
Help Topics	F1
Exit	Alt+F4

10.3. Field Verification Window

Confirm Field	Enter
Postpone	Ctrl+Enter
Next Uncertain Character	F4
Previous Uncertain Character	Shift+F4
Next Field	Page Down
Previous Field	Page Up
Undo	Ctrl+Z
Redo	Ctrl+Y
Cut	Ctrl+X
Copy	Ctrl+C
Paste	Ctrl+V
Delete	Del
Delete All	Alt+Del
Select All	Ctrl+A
Insert Line Break	Shift+Enter
Merge Characters	Ctrl+M
Analogous Fields	Alt+F3
Show Character Image	F2
Full Screen	F11
Field Data: Recognized Text	Alt+F1
Field Data: Character Images Cutting	Alt+F2

Show Field Image	Ctrl+I
Field Image: On Top	Alt+1
Field Image: On Bottom	Alt+2
Zoom In	Ctrl+Num+
Zoom Out	Ctrl+Num–
Help Topics	F1
Exit	Alt+F4