

Automated Key-Value Extraction with the Document Labels (KVP) Model in Scan2x



1. Use Case Overview



Goal

To automatically ingest structured or semi-structured documents (e.g., forms, questionnaires, order sheets), capture key-value pairs via the Document Labels (KVP) model, handle handwritten text, and map these dynamic fields to normalised metadata defined in Scan2x for seamless integration with downstream systems.



Primary Actor

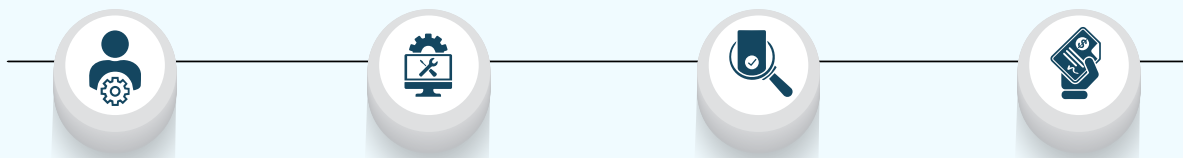
- Data Entry Specialist or Document Processing Operator



Supporting Actors

- **Scanning Hardware** (if physical documents are scanned)
- **Downstream Systems** (CRM, ERP, compliance databases, or other business platforms)

2. Stakeholders and Interests



Operations / Administrative Teams

- Want to reduce the time spent manually entering data from a variety of forms and documents.
- Need a flexible solution that can handle diverse document formats with minimal setup.

IT Department

- Seeks an easily maintainable, minimal-configuration solution that can adapt to changing document formats.
- Ensures integration with enterprise systems and data security.

Compliance / Auditing Teams

- Requires clear audit trails of how data is extracted and normalised.
- Needs consistent, accurate data from critical documents.

Finance / HR / Other Departments

- Beneficiaries of accurate, digitised data for their own workflows (e.g., purchase orders, HR forms).

3. Preconditions



System Configuration

- Scan2x is installed, configured, and accessible to authorised users.
- The Document Labels (KVP) model is enabled in Scan2x.



Document Handling

- Documents can be paper-based, scanned images, PDFs, or digital forms.
- Handwritten data in fields is acceptable; the model's handwriting recognition capability is turned on.



Language Support

- According to Scan2x documentation, the Document Labels (KVP) model supports multiple languages (e.g., English, Spanish, French, German, and others).

- Users can configure the system to process these languages as required by their documents.



Document Handling

- Documents can be paper-based, scanned images, PDFs, or digital forms.
- Handwritten data in fields is acceptable; the model’s handwriting recognition capability is turned on.

4. Triggers



Manual Trigger

- An Operator scans or uploads files (PDF, image, etc.) to Scan2x.



Automated Trigger

- Documents arrive in a monitored folder or email inbox, and Scan2x automatically processes them using the Document Labels (KVP) model.

5. Main Success Scenario (Basic Flow)

1

Document Ingestion

- The user or the system uploads/receives a batch of documents.
- These documents may contain typed text, tables, or handwritten entries in specific fields.

2

Document Classification

- Scan2x identifies the document type (e.g., “Order Form,” “HR Onboarding Form,” “Insurance Claim”) by layout, barcodes, or text patterns.
- The Document Labels (KVP) model is triggered for recognised document types.

3

Key-Value Pair Extraction

- Scan2x uses its Document Labels (KVP) model to locate and extract label-value pairs. Examples:
 - **Field Label:** “Full Name” → **Value:** “Jane Doe”
 - **Field Label:** “Amount” → **Value:** “\$1,250.00”
 - **Field Label:** “Signature” → **Value:** [Handwritten name or signature block]
- Handwriting recognition is applied if the values are handwritten, seamlessly converting cursive or printed handwriting into digital text.

4

Language-Agnostic Processing

- If the document is in a supported language, the system automatically extracts the relevant key-value pairs in that language.
- The user can select or configure the language if necessary, ensuring robust multilingual support.

5

Mapping Labels to Normalised Fields

- Although the extracted key names may vary per document (e.g., “Client Name,” “Customer Name,” “Name”), within Scan2x these labels can be mapped to a normalised field name (e.g., “CustomerFullName”) defined in the system.
- The Operator or Admin has configured these mappings once, so subsequent documents automatically follow the same mapping.

6

Validation

- The system flags any potential inconsistencies or missing data (e.g., a required field is unreadable).
- An Operator can review flagged items and correct them within Scan2x.

7

Data Export

- Once validated and mapped, the normalised data is exported or posted to the downstream system, or saved in a standard file format (e.g., CSV for a spreadsheet, JSON for an API).
- A success message or reference ID returns to Scan2x, confirming completion.

8

User Notification

- The Operator (and optionally others) receives confirmation that the key-value extraction and mapping process was successful.
- Documents are stored or archived within Scan2x, and the normalised metadata is now ready for reporting or further processing.

6. Alternative Flows and Exceptions



Incomplete Labels or Unrecognised Fields (Alternative Flow)

- **Trigger:** Document contains unusual or non-standard field labels.
- **Action:** System flags the unrecognised label for review. An Operator defines a new mapping, which Scan2x then uses for future documents.



Poor Handwriting or Low-Quality Scans (Alternative Flow)

- **Trigger:** Handwriting is illegible or the scan is distorted.
- **Action:** Scan2x either attempts to interpret the text or flags it for manual review. The user can re-scan or correct the data manually.



Language Identification Issue (Alternative Flow)

- **Trigger:** The document language is not clearly detected.
- **Action:** User specifies the document’s language, or Scan2x attempts next-best recognition. Once set, extraction proceeds normally.



Export Failure (Exception)

- **Trigger:** Network or API error prevents export of the data.
- **Action:** The system retries at intervals or notifies the Operator. Data is queued until the connection is restored or a manual export is performed.

7. Postconditions



Successful Extraction and Mapping

- The system has extracted key-value pairs (including handwritten text) and automatically normalised them via label-to-field mapping.
- Downstream systems receive consistent data as required.



Error Handling and Notifications

- Any documents with incomplete or invalid data remain in a review queue until resolved.
- Operators and admins have full visibility into any flagged items.

8. Benefits and Outcomes



Minimal Configuration: The Document Labels (KVP) model can handle a range of structured and semi-structured documents with minimal setup, only requiring an initial mapping of label names.



Handwriting Recognition: Built-in handwriting recognition extracts even freeform or cursive entries, reducing the need for manual transcription.



Multilingual Support: Scan2x can process documents in multiple languages, broadening its applicability across global organisations.



Accurate Normalisation: Despite varying label names in source documents, Scan2x maps them consistently to defined normalised fields, ensuring reliable integration with downstream systems.








Increased Efficiency: Automates what would otherwise be time-consuming manual data entry, freeing staff for higher-value tasks.

Use Case Summary

By deploying the Document Labels (KVP) model within Scan2x, organisations can rapidly ingest and process diverse types of structured and semi-structured documents—extracting both printed and handwritten key-value pairs. This minimal-configuration approach allows businesses to quickly adapt to new document layouts, leveraging powerful multilingual and handwriting recognition capabilities. The dynamic labels found within each document are then mapped to standardized fields in Scan2x, ensuring a seamless flow of consistent, normalized data into any required downstream system.



 **Avantech Ltd** avantechsoftware.com
 +356 2148 8800
 software@avantechsoftware.com
 Avantech Ltd |  @Avantechsoftware