NLQ-Enabled Document Analysis with Scan2x

What is NLQ (Natural Language Query)?

1. Use Case Overview

Goal
To ingest a wide range of document types (e.g., PDF, image, scanned forms), submit them to Scan2x for NLQ Processing, and retrieve specified fields or summaries in a structured format—based on user-defined descriptions and contextual prompts.

Primary Actor
Document Processing Specialist or Business User.

Supporting Actors
Scan2x Platform (configured for NLQ processing)
Downstream Systems (databases, ERPs, analytics tools) that receive the extracted data

2. Stakeholders and Interests

Business Operations / Admin Staff
- Want to extract specific information from complex or varied documents without heavy configuration.
- Need to handle a broad variety of document layouts or content structures seamlessly.

IT Department / Scan2x Administrators
- Configure the NLQ settings, including prompts and field definitions.
- Ensure data security and integration with other enterprise systems.

Legal / Compliance Teams
- May need deeper insights from policies, contracts, or regulatory documents to ensure compliance and manage risk.

Data Analysts / Researchers
- Use the extracted data for reporting, analytics, or further review.
- Require context-based extraction for nuanced details not easily captured by simpler OCR templates.

3. Preconditions

Scan2x Configuration
- Scan2x is installed and accessible to authorised users.
- The Processing with Natural Language Query feature is enabled.

Prompt and Context Setup
- Administrators define a general prompt or context about the documents to be analysed (e.g., “These are marketing disclaimers,” or “These are internal policy documents”).
- Administrators also create normalised field definitions with plain-language descriptions (e.g., “Author Name,” “Document Summary,” “Key Dates”).

NLQ Integration
- Scan2x includes advanced language and vision-based analysis capabilities to interpret document text, images, and layouts.

User Access
- Relevant staff have permissions to upload documents and initiate NLQ processing in Scan2x.

4. Triggers

Manual Document Submission
- A user uploads or scans a document into Scan2x and selects “NLQ Processing” as the desired method of analysis.

Automated Queue
- Documents placed in a monitored folder or email inbox can be automatically routed to the NLQ workflow in Scan2x.

5. Main Success Scenario (Basic Flow)

Document Ingestion
- The user or system adds a document (PDF, image, scan) to Scan2x.
- Scan2x classifies the document type and applies an NLQ analysis workflow as configured.

Field Descriptions and Requirements
- Each normalised field is accompanied by a plain-language description explaining exactly what data is needed.

Example:
- **Field:** “Key Risks”
- **Description:** “Identify any statements in the document mentioning potential dangers, hazards, or disclaimers.”

Structured Output
- For each predefined field, Scan2x returns the information discovered in the document.

Example:
- **“Key Risks”:** “Potential supply chain delays due to recent policy changes.”
- **“Action Items”:** “Schedule a review meeting before March 1.”

Normalisation and Mapping
- The extracted text is aligned with the standard field structure in Scan2x for consistent output across all documents.
- If external systems require specific naming conventions, the data can be mapped accordingly.

User Notification
- The user receives a success message.
- The document and extracted metadata are stored and available for further business processes in Scan2x.

1

Context Prompt Application
- Scan2x includes the administrator-defined context (e.g., “Analyse the following document for marketing disclaimers...”) to guide the analysis.
- This context helps focus on relevant aspects of the document.

2

3

NLQ Analysis
- Scan2x processes the document’s text and visuals, interpreting each section according to the prompt and field descriptions.
- It extracts or summarizes the relevant sections, text blocks, or data points.

4

5

Validation and Review
- The results are displayed for user verification in Scan2x.
- The user can accept, refine, or add more context if needed.

6

7

Export / Integration
- The validated data is exported to a downstream system (e.g., CRM, compliance tool) or saved in formats like CSV or JSON.
- Scan2x logs confirmation and references for audit purposes.

8

9

6. Alternative Flows and Exceptions

Ambiguous or Incomplete Document
- **Trigger:** The document lacks clear text or is partially unreadable.
- **Action:** Scan2x flags the data for review or requests additional user input to clarify the missing details.

Inconsistent Context
- **Trigger:** The context provided doesn't match the actual document type.
- **Action:** The user updates the prompt or reclassifies the document, then re-runs the NLQ process.

Network / System Errors
- **Trigger:** Technical issues prevent analysis from completing.
- **Action:** Scan2x retries or notifies the user to correct the error (e.g., network outage, file corruption).

High-Sensitivity Documents
- **Trigger:** The document contains confidential information.
- **Action:** Security settings in Scan2x either block unauthorised processing or anonymise sensitive fields before NLQ processing.

7. Postconditions

Successful Analysis
- Scan2x has extracted targeted text for each normalised field, guided by the plain-language prompts and context.
- Data is available in a structured format for immediate use or downstream consumption.

Audit Trails and Review
- Scan2x maintains a history of all requests, outputs, and user edits.
- Administrators can review these logs to refine the prompts or address any compliance considerations.

8. Benefits and Outcomes

- **Flexible, Context-Based Extraction**
  - NLQ allows administrators and users to define precisely what they want from a document, using everyday language and a general or specific context prompt.

- **Minimal Configuration**
  - Users can adapt to new or varied document types simply by adjusting plain-language field definitions and context.

- **Reduced Manual Labor**
  - Automated interpretation of documents—even complex or lengthy ones—significantly cuts the time needed for data entry and review.

- **Consistent, Normalised Results**
  - Despite varied document formats, the final extracted data is mapped into standard fields for easy integration with other systems.

- **Improved Decision-Making**
  - Relevant insights are quickly accessible, supporting faster, data-driven decisions across different organisational functions.

Use Case Summary

Using Scan2x Natural Language Query Processing, organisations can swiftly interpret and extract vital information from documents of virtually any format. By providing both a contextual prompt and plain-language field definitions, Scan2x guides the analysis process, returning structured, normalised data ready for downstream systems and decision-making. This approach reduces setup time, increases accuracy, and accommodates a wide array of document types—all without requiring specialised template creation or extensive configuration.

Avantech Ltd avantechsoftware.com
+356 2148 8800 | +44 780-9757598
software@avantechsoftware.com
Avantech Ltd | @Avantechsoftware