



AutoContract.AI



Problem Statement

Managing and processing lengthy, intricate contract documents presents significant challenges for organizations. These documents, often stored in varying formats like PDF, DOCX, or TXT, require accurate extraction of critical information, verification for compliance, and structured output generation. Manual processing is labor-intensive, prone to errors, and lacks scalability, while ensuring compliance and managing version control adds complexity.

Key Challenges

- **Document Complexity:** Contracts feature dense legal language, varied formats, and inconsistent structures, making automated processing difficult.
- **Scalability:** Managing high volumes of documents manually is inefficient and resource-intensive.
- **Compliance Assurance:** Ensuring all extracted data adheres to legal and business rules requires robust validation mechanisms.
- **Version Control:** Tracking and comparing document amendments for accuracy and traceability is challenging.
- **Efficient Search and Retrieval:** Quickly locating and linking related contracts, clauses, and amendments is critical for operational efficiency.

Objective

The objective is to automate the handling of contract documents, reducing manual effort and errors while improving accuracy in data extraction and compliance validation. By leveraging SAP Gen-AI and SAP HANA, the solution ensures efficient processing and retrieval of high volumes of contracts through advanced similarity search and retrieval capabilities.

Additionally, the solution aims to provide an intuitive user experience through a chatbot interface for document management, enabling seamless uploads, reviews, and structured output generation. Robust version control and compliance tracking mechanisms ensure legal and operational standards are upheld while maintaining traceability for all contract amendments and updates. This empowers organizations to make faster, data-driven decisions with confidence.

Scope

The solution automates the processing of lengthy contract documents in formats such as PDF, DOCX, and TXT by extracting structured information through predefined questions, semantic analysis and generative AI. It ensures accuracy in data extraction while automating compliance verification against predefined legal and business rules to highlight potential discrepancies. Additionally, the system supports document versioning and amendment tracking, enabling users to compare changes and maintain a clear history of revisions.

Metadata extraction and enrichment enhance indexing and search capabilities, allowing efficient retrieval of related documents through SAP HANA's vector-based similarity search. Structured outputs, such as Excel templates, are generated for user review and decision-making. These features integrate seamlessly with SAP tools, ensuring a streamlined workflow for contract management.

Out of Scope

- Changes or updates in compliance rules and validation, during and after development, not limited to legal or internal process within company, are not accommodated for during the development of the solution.
- Automated/Intelligent Document Redaction for redacting sensitive information is not included as a part of data processing pipeline or within response.
- Integration with SAP and third party systems is not included within the scope of this POC project.
- The solution is limited to processing Contract Type Documents, other documents are not considered within scope
- Queries for the Generative AI solution marked as out of scope by company are excluded during development and internal testing
- Any additional questions apart from those listed in requirements, will be considered out of scope

Key Assumptions

- The quality of the responses is dependent upon the image/text quality of the documents
- Team shall identify SPOC for functional/technical clarifications who would be empowered to make decisions on behalf of the function they represent. SPOC would be made available as per the detailed project plan that will be developed and shared during the Project Preparation phase for this engagement.
- UAT is the responsibility of the Lemark team. All necessary personnel and resources to execute activities associated with testing will be the responsibility of the client.
- All compliance and validation rules must be explicitly stated in the requirements
- The solution is optimized for the accuracy of the response.
- The solution will primarily process documents in English and all estimates are provided based on this assumption. Additional requirements with regards to languages must be specified and will impact the development process.
- Crave InfoTech and company will work together to prepare mapping document for source and target system.
- Team to provide the access to SAP services and S/4HANA during the development phase. If it is not available, then Crave will use own BTP tenant for development
- Source and target systems are made available for development and testing.

Approach: Document Management and Analysis Process

Input Management

- **User Interface Screen:** Users begin by uploading documents to the system.
- **Document Parsing:** The uploaded document is parsed into manageable pieces.
- **Text Chunking:** The parsed data is then segmented into chunks suitable for processing.

Data Processing

- **Text Embedding Generation:** Converts text chunks into numerical data that can be easily processed by machine learning algorithms.
- **Embedding Storage:** Stores the generated embeddings for further use in the process.

Information Extraction

- **Predefined Questions:** Uses set questions to extract specific information from text data.
- **Model Selection and Training:** Selects and trains models based on the embeddings to improve accuracy of information extraction.

Compliance Verification

- **Rule-Based Compliance Checks:** Ensures data meets specified regulatory standards.
- **LLM-Based Consistency Verification:** Uses large language models to verify the consistency of the extracted data.

Data Output

- **Excel Report Generation:** Compiles the extracted information into Excel format for easy analysis.
- **Output Delivery:** Delivers the final processed data to the end user.

Approach: In-Depth Analysis of Key Process Components

Enhanced Document Management

- **Version Control Integration:** Keeps track of different versions of documents for auditing and historical data analysis.
- **Amendment Extraction and Compliance:** Detects and extracts amendments made to documents, ensuring all changes comply with standards.

Metadata Management

- **Metadata Extraction:** Pulls relevant metadata from documents.
- **Metadata Storage and Enrichment:** Enhances the metadata by adding more context or correcting existing information for better usage.

Chatbot Integration

- **Chatbot Model Creation:** Develops a tailored chatbot model that uses trained data to interact effectively with users.
- **Training:** Regular updates and training cycles improve the chatbot's responses and functionality.
- **Provide Insights:** The chatbot quickly provides insights based on user queries, enhancing user interaction with the system.

Output and Review Process

- **Layout-Aware Prompting for Documents:** Adjusts the document layout during the output phase to enhance readability and presentation.
- **User Interaction and Review:** Allows users to interact with the final outputs and provide feedback, which is crucial for continuous improvement.
- **Downloads:** Facilitates the downloading of reports and documents in user-preferred formats.



**Your Success is Our Priority
Lets Win Together!**



+1 844 CRAVEIT (272-8348)



contact@craveinfotech.com



www.craveinfotech.com



15 Corporate Pl S #104, Piscataway, NJ 08854

