IDA Understanding

Understanding documents with automatic insights

IDA Understanding doesn't just extract data, it comprehends documents. The module uses Large Language Models to answer questions, create summaries, and analyze content that isn't explicitly stated in the text. Unlike chat applications, IDA Understanding works through batch processing and writes insights directly into document metadata. For seamless integration into downstream systems like Enterprise Content Management, legal databases, or compliance tools..

KEY FEATURES

Prompting in natural language

Define your analysis requirements in natural language without programming and without complex query syntax. IDA Understanding automatically executes a list of recurring prompts for all input documents and writes results directly into document metadata.

AI agent for tables of contents

An integrated AI agent automatically generates tables of contents, aiding users in the structured organization of lengthy documents. The agent initially divides the document into manageable sections to create concise content summaries for each page. Subsequently, it compiles a table of contents for the entire document.

Easy deployment and integration

IDA is deployed either on-premises or in a (private) cloud as a Java application or containerization using Docker.

The gRPC API (alternatively REST API) facilitates swift integration.

REQUIREMENTS

The **IDA Server** is required to process input documents and provides a browser interface.

For 64-bit systems

Linux: Ubuntu 18.04-25.10, Debian 11-13, CentOS 8-10, Red Hat 8.x-10.x, LEAP 15.4-15.6, 16.0; SLES 15 SP 4-7

Windows: 10, 11

Windows Server: 2016, 2019, 2022, 2025

Docker

At least **12 GB hard disk storage**

At least **16 GB RAM**

Leveraging unmatched OCR quality

IDA Understanding is built on **IDA Recognition**, the OCR engine for outstanding results in the most difficult scenarios. Even with distorted scans, poor image quality, and difficult-to-read handwriting, IDA Recognition delivers a high-quality text foundation. Why this matters: Output quality depends entirely on input data quality. IDA Recognition captures machine-printed and handwritten text, checkboxes, tables, and historical scripts as the perfect foundation for downstream processes.

Batch processing instead of chat interface

IDA Understanding is not a chat application but designed for mass processing.

Chat applications (like JAIDE)

- Interactive individual queries
- Manual document uploads

IDA Understanding

- Automated batch processing
- Results as structured metadata

How does it work?

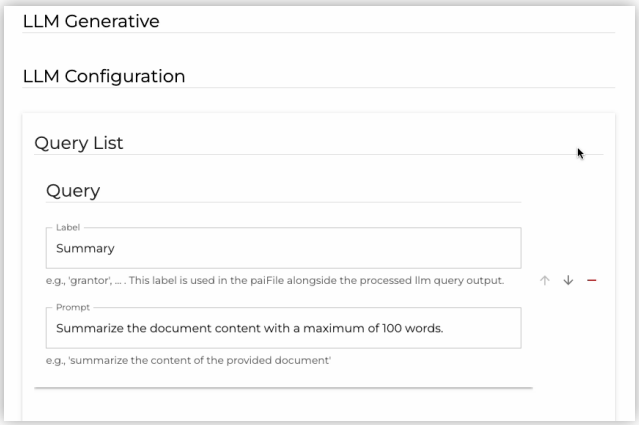
IDA Understanding is **best suited for handling unstructured documents** that lack fixed layouts or data points, such as contracts or cover letters.

A typical use case is **legal discovery**: IDA Understanding automatically reviews large file volumes for relevant facts, identifies critical clauses and risks, and generates summaries for quick decision-making.

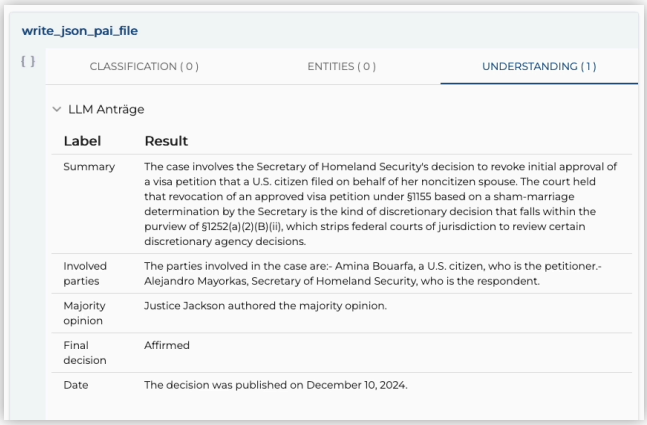
The difference to IDA Extraction:
While LLM Entity Extraction finds and marks explicitly present information (“Where is the amount stated?”), IDA Understanding generates insights from the overall context (“What is the main message of this document?”). IDA Understanding is particularly useful when content isn’t explicitly stated in documents, whereas LLM Entity Extraction focuses on finding entities and highlighting them in their original positions.

Use both modules together for maximum efficiency: Entity Extraction for structured data capture, Understanding for context-based analysis and insight generation.

Technical foundation: Like LLM Entity Extraction, IDA Understanding uses large language models. However, instead of extracting data based on keywords, it extracts or generates information based on prompts.



Input: List of prompts for court opinions



Output: Metadata in IDA Web Client

LLM SERVER

To utilize large language models (LLMs) **on-premises**, a dedicated **LLM Server** is necessary:

For 64-bit systems

- **Docker (Ubuntu-based)**
- At least **40 GB GPU memory** (can be spread out across multiple GPUs)
- At least **6.5 GB hard disk storage** + at least 20 GB for LLM
- At least **64 GB RAM**

Important: Disk storage and RAM depend heavily on the chosen models. Note that a CPU-only mode is not possible.

The LLM Server can also connect to **OpenAI models** and then acts as a relay. This means that no GPUs are necessary.

Experience IDA Understanding for yourself and contact us for a personalized demo.

Book demo