

IDA Understanding

Actionable insights from unstructured documents

IDA Understanding is designed to **automatically deliver actionable insights from unstructured documents**, such as generating summaries and performing question-answering. Unlike a chat application, these insights are produced through mass processing and can be seamlessly integrated into subsequent processes and applications, such as enterprise content management.

KEY FEATURES

Prompting in natural language

IDA Understanding automatically executes a list of prompts on input documents, writing answers in the documents' metadata. These prompts can be written in natural language and can include summaries or questions.

Automatic table of contents

IDA Understanding includes an **AI agent** for the automatic generation of tables of contents, aiding users in the structured organization of lengthy documents. In collaboration with a Large Language Model (LLM), 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.

SYSTEM 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, 12; CentOS 8, Red Hat 8.x, 9; LEAP 15.x, SLES 15 SP 4-6

Windows: 10, 11

Windows Server: 2016, 2019, 2022

Docker

At least **12 GB hard disk storage**

At least **16 GB RAM**

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

Leveraging unmatched OCR quality

IDA Understanding is based on [IDA Recognition](#), an optical (OCR) and intelligent (ICR) character recognition engine that delivers outstanding results even when dealing with the most difficult scenarios. IDA Recognition captures machine-printed and handwritten text, checkboxes, tables, and historical scripts, even in poor-quality scans with rotated or skewed print. Having high-quality input data is crucial for subsequent tasks as it directly affects output quality.

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.

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.

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.

SYSTEM REQUIREMENTS

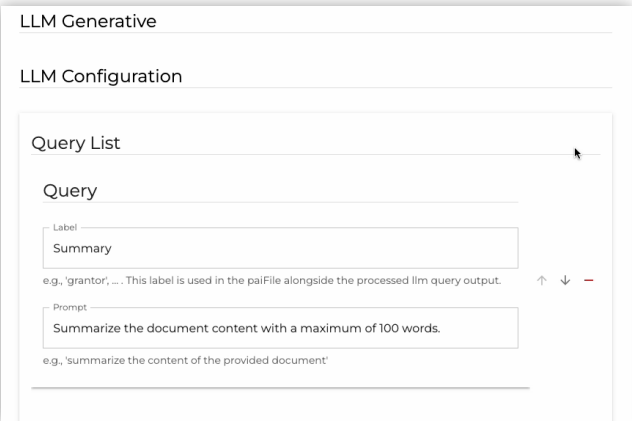
To utilize large language models (LLMs) on-premises, a dedicated server, referred to as the **LLM Server**, is necessary:

For 64-bit systems

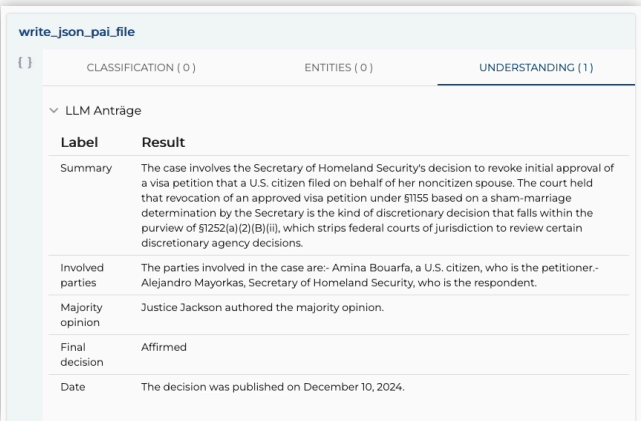
- **Docker (Ubuntu-based)**
- At least **40 GB GPU memory** (can be spread out across multiple GPUs)
- At least **6.5 GB hard disk storage** + at least 20 GB for LLM
- At least **64 GB RAM**

The required hard disk storage and the necessary RAM depend significantly on the models intended to run on the LLM Server. Note that a CPU-only mode is not possible.

The LLM Server can also connect to **OpenAI models**, which can make extensive hardware setups redundant.



Input: List of prompts for court opinions



Output: Metadata in IDA Web Client

For more information, please refer to the [software documentation](#).