Structured Data Extraction

LLMs are capable of ingesting large amounts of unstructured data and returning it in structured formats, and LlamaIndex is set up to make this easy.

Using LlamaIndex, you can get an LLM to read natural language and identify semantically important details such as names, dates, addresses, and figures, and return them in a consistent structured format regardless of the source format.

This can be especially useful when you have unstructured source material like chat logs and conversation transcripts.

Once you have structured data you can send them to a database, or you can parse structured outputs in code to automate workflows.

Full tutorial

Our Learn section has a full tutorial on structured data extraction. We recommend starting out there.

There is also an example notebook demonstrating some of the techniques from the tutorial.

Other Guides

For a more comprehensive overview of structured data extraction with LlamaIndex, including lower-level modules, check out the following guides:

- Structured Outputs
- Pydantic Programs
- Output Parsing

We also have multi-modal structured data extraction. Check it out.

>> Ask AI

Miscellaneous Examples

Some additional examples highlighting use cases:

- Extracting names and locations from descriptions of people
- Extracting album data from music reviews
- Extracting information from emails

Ask AI