

Extraction

Conceptual Guide

Most APIs and databases still deal with structured information. Therefore, in order to better work with those, it can be useful to extract structured information from text. Examples of this include:

- Extracting a structured row to insert into a database from a sentence
- Extracting multiple rows to insert into a database from a long document
- Extracting the correct API parameters from a user query

This work is extremely related to [output parsing](#). Output parsers are responsible for instructing the LLM to respond in a specific format. In this case, the output parsers specify the format of the data you would like to extract from the document. Then, in addition to the output format instructions, the prompt should also contain the data you would like to extract information from.

While normal output parsers are good enough for basic structuring of response data, when doing extraction you often want to extract more complicated or nested structures. For a deep dive on extraction, we recommend checking out [kor](#), a library that uses the existing LangChain chain and OutputParser abstractions but deep dives on allowing extraction of more complicated schemas.