

## ThoughtRiver - Document Headers

Legal agreements are typically quite structured documents. They usually open with a title and a declaration of the parties to the agreement. This is often followed by a declaration section where key terms are defined to disambiguate later clauses.

The rest of the document is the body of the contract (which is often broken into sections) followed by a brief signature (“execution”) block.

---

The goal of this task is to produce an installable package that can extract headers from a document. The implementation should be object-oriented and take the provided JSON format as input. Please build the package as you would if it were to go into production. It should provide an interface (with suitable methods or functions) to retrieve the headers, given the document as an input. The case where no headers are present should be handled appropriately. The package should be built such that it is easy to add additional logic and should be covered by appropriate tests.

The actual logic to determine whether a paragraph is a header does not need to be 100% accurate, but we are interested in your ideas and the implementation of them. This task is designed to examine the underlying ideas together with the design, architecture and implementation of them and the supporting package.

Please provide a supporting document detailing your solution, how you came to the solution and any known limitations with it.

For reference, I have included two documents in a JSON format and the source file that they were derived from. Each row in the JSON corresponds to a paragraph in the document. Additionally, I have included `NDA1.json` in an HTML format where the headers are highlighted.