

MCP サーバから LLM を使う Sampling 機能を試してみる

大野 遼太

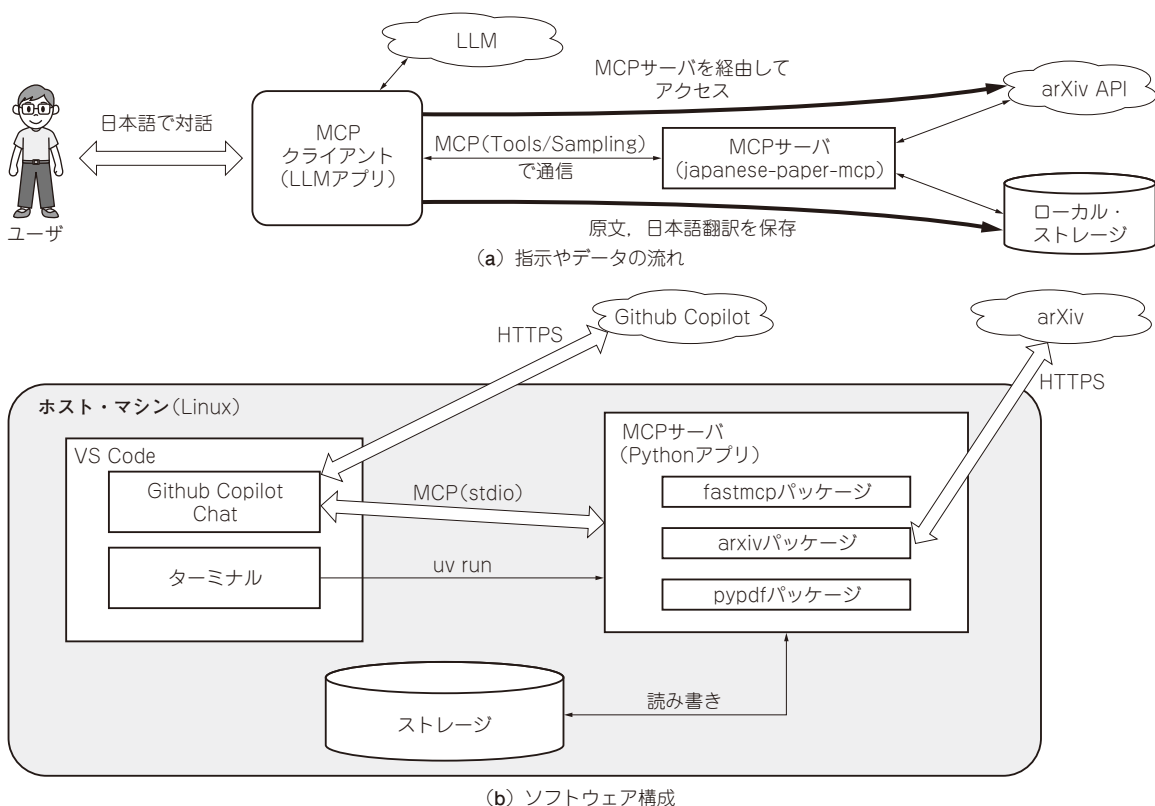


図1 作成する MCP サーバの Sampling 機能を使った論文検索システム

トライすること

本稿では、MCP (Model Context Protocol) の Sampling 機能を利用した独自の MCP サーバ (japanese-paper-mcp MCP サーバ) を作成します。構成を図1(a)に示します。Sampling 機能とは、MCP サーバがクライアント (AI エージェント) を介して LLM (Large Language Models) の推論 (結果生成や補完) を実行する機能を指します。通常であれば LLM の推論はクライアント側で行われますが、この Sampling 機能を利用することで、MCP サーバ側から

クライアントに LLM による推論をリクエストできます。

ポイントとなるのは、LLM と接続しているのは MCP クライアントのみで、MCP サーバは接続していないということです。MCP サーバが直接 LLM とやり取りすることがないので、MCP サーバ側に API キーなどを渡す必要はありません。

● 英語資料の検索&翻訳エージェントを作る

今回は例として、MCP サーバ、MCP クライアントを使って、LLM と外部のサーバとをつなぎ、日本語で指定した興味あるテーマの英語資料を LLM が適切に外部サーバから検索し、内容を日本語に翻訳してく

◆参考文献◆

- (1) About MCP Understanding MCP clients. <https://modelcontextprotocol.io/docs/learn/client-concepts#sampling>