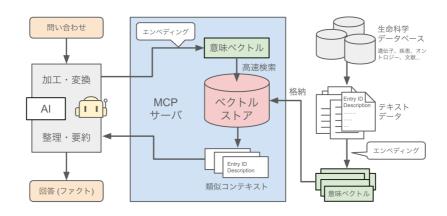
統合データベース検索におけるテキストエンベディング利活用



千葉啓和、藤原豊史、守屋勇樹、池田秀也、申在紋 (ライフサイエンス統合データベースセンター)

Methods

バイオメディカルDBのエントリー毎に、テキストをEmbedding化することでフレキシブルな検索を実現



ライフサイエンス統合データベースセンター (DBCLS) では、大規模言語モデル及びエンベディング技術を用いて、データベースと非構造化データを統合的に利用するための技術開発を進めている。 特にここでは、テキストエンベディングを活用してデータベースを検索するシステムの開発について発表する。本システムでは、バイオメディカル分野のデータベースに含まれるエントリーを対象 として、事前にテキストエンベディングを実行し、各エントリーの埋め込み表現を生成する仕組みを構築している。さらに、整備済みのデータセットに対して、ユーザーが入力した自然文の埋め込 み表現に基づくベクトル検索を行い、検索結果を要約して提示することができる。検索結果を要約する際に、データベースエントリーをファクトとして付加することにより、ユーザーが入力した文章に関するファクトチェックを支援することもできると考えている。本発表では、開発中のシステムについて紹介したい。

Results

バイオメディカルDBに対するフレキシブルな検索を実現し、ファクトチェックやアノテーションに応用

