

Notionデータを対象としたベクトル検索システム 「NotionFinder」の開発

龍谷大学 先端理工学部 数理・情報科学課程

Y220010 田代 圭梧

指導教員 佐野 彰

概要

近年、情報管理ツールとしてNotionの利用が急増しているが、蓄積された膨大なデータに対する検索機能はキーワードマッチングに依存しており、文脈や意図を考慮した検索が困難であるという課題がある。また、公式のNotion AIはコストや外部公開時の制約から、Webサイトとして構築されたナレッジベースでの利用には適さない場合が多い。

本研究では、Notion内の非構造化データを対象とし、文脈に基づいた検索（セマンティック検索）を可能にするWebプラットフォーム「NotionFinder」を開発した。

本システムは、Notion APIを用いて階層的なブロックデータを再帰的に取得し、軽量の埋め込みモデル（multilingual-e5-small）を用いてベクトル化を行う前処理パイプラインと、Next.jsおよびHNSWアルゴリズムを用いた高速な検索実行環境から構成される。特に、サーバーレス環境（AWS Amplify Hosting）におけるリソース制約下での運用を想定し、Transformers.jsを用いたサーバーサイド推論や、インデックスの事前構築手法を取り入れた。なお、埋め込みモデルの選定および検索パラメータの決定は、検索精度と検索速度のトレードオフを評価する実験に基づいて行った。

検索精度に対する評価実験の結果、キーワードが完全一致しない抽象的なクエリに対しても、意図に合致するドキュメントを高い精度で抽出できることを確認した。さらに、検索速度に対する評価では、ウォームスタート時において約0.5秒という実用的な応答時間で結果を返却できることを確認した。

本研究では、高価な外部サービスに依存することなく個人のNotionデータを安価かつ高速に検索できるWebアプリケーションとして NotionFinder を実装し、評価実験によりその有効性を確認した。