

研究成果1: 構造化RAGのためのJSONLの高速部分構造検索

構造化データ検索を通じて、LLMのRAG・プロンプト検索の基盤技術としての応用を想定。

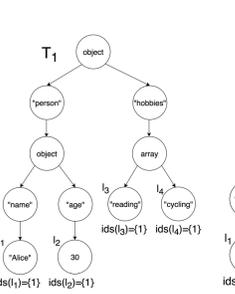
背景: JSONLは、LLMプロンプト、化合物、地理空間データなどの保存形式として広く普及しており、構造化フォーマットを用いることで推論精度が最大40%向上することが報告されている。

既存手法の問題点: 従来の木探索やJSON→XML変換に基づく手法では、クエリごとにデータ全体を走査する必要があり、計算量がデータセットサイズに対して線形に増大するため、大規模JSONLデータに対しては実用的ではない。

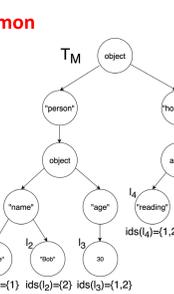
解決策: 共通の構造パスを統合する「マージ済み木」と、木構造を線形化して高速なナビゲーションを可能にする圧縮データ構造XBWを組み合わせた簡潔索引jXBWを提案。さらに、クエリをパスに分解する3段階アルゴリズムにより、データ規模ではなくクエリの構造に依存する計算量で部分構造検索を実現。

結果: 実験では、従来の木ベース手法に対して最大4,700倍、XMLベース手法に対しては6×10⁶倍以上の高速化を達成した。大規模化合物データ上でjXBWによるRAGでLLMが創薬候補化合物を生成できるか簡易的に確認。

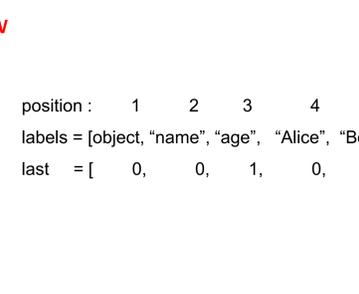
(i) JSON Tree T1, T2



(ii) Merged Tree



(iii) XBW representation of MT



jXBW achieves consistent sub-millisecond query times



研究成果2: Dynamic r-index: An Updatable Self-index in LCP-bounded Time, DCC2026に採択

本研究は、大規模反復文字列に対して初めて再構築不要で実用的な動的self-indexを実現した。

背景: ゲノムなど反復の多い文字列に対し、r-indexは空間効率と高速検索を両立する標準手法

問題点: r-indexは静的構造のため、更新のたびに再構築が必要

提案手法: r-indexを構成するrun-length BWTと、run境界でサンプリングした接尾辞配列を動的管理し、LCP配列に基づき更新影響を局所化

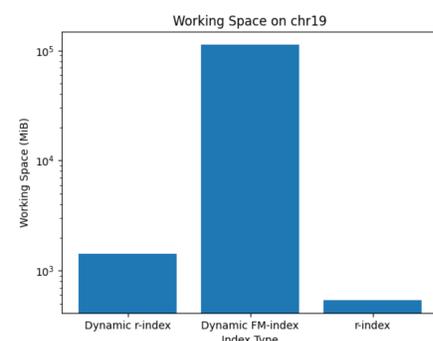
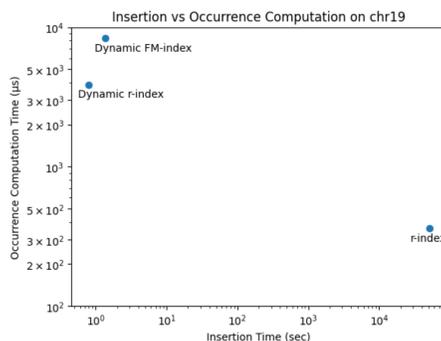
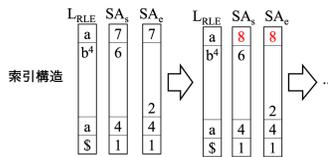
結果: 挿入・削除をO((m + Lmax) log n) (Lmax: LCP最大値) で実現

実験: 1000人分ヒトゲノム(約59GB)に対し1文字挿入あたり約0.8-0.9秒で更新可能。再構築が必要な静的r-index(数時間~十数時間)より桁違いに高速。検索速度は実用的な時間、動的FM-indexより大幅に省メモリ

Run-length BWTの更新に従って変化する部分のみ修正

文字列 (bを挿入): b b a b b a S

Run-length BWT: ab⁷a⁷S



公開ソフトウェア: Deep OpenReview Research — 国際会議論文の効率的な発見・分析エージェント

研究者がAI分野の大型国際会議(例: NeurIPS, ICLR, ICML)で採択された数千件の論文の中から、自分の研究興味に合致する重要論文を効率的に発見・分析するためのツールを開発・公開した。

背景: NeurIPSやICMLでは毎年数千本の論文が採択され、ユーザーの興味に一致する論文を発見することが困難

問題点: 従来の検索はキーワードの完全一致ベースであり、同義語・表記揺れを考慮できない。レビュー情報を活用できないため、採択理由や新規性・実用性評価といった深い理解は不可能。結果として関連論文の発見・優先順位付けが非効率である。

解決策: (i) LLMとOpenReviewデータを用いた意味理解ベースの論文探索・評価; (ii) 同義語展開による意味検索; (iv) Meta Review・Decisionを含むレビュー解析; (v) 関連性・新規性・実用性などの多軸評価; (iv) 自然言語入力による検索指定; (vi) Markdown形式での自動レポート生成

結果: Deep OpenReview Researchを用いることで、従来のキーワード検索では見逃していた関連論文を効率的に発見・多軸評価し、ランキング付きレポートとして出力できるようになった。たとえば、数千件から上位100件を抽出し、関連性・新規性・インパクトなどの観点で順位付けしたレポートを生成することができる。

日本語版GitHubへのリンク: <https://github.com/tb-yasu/deep-openreview-research-ja>

英語版GitHubへのリンク: <https://github.com/tb-yasu/deep-openreview-research>



公開約1ヶ月で累計400クローン達成

