

PostgreSQLでの セマンティックサーチへの挑戦

株式会社クリアコード
堀本 泰弘/阿部 智晃

PostgreSQL Conference Japan 2025
2025-11-21

本日の内容

1. キーワード検索の課題と改善の方向性
2. セマンティックサーチを実現するために
3. RaBitQを採用した理由
4. PostgreSQLのセマンティックサーチ
5. 性能
6. 用例：FAQ検索・ナレッジベース検索の高度化

キーワード検索の課題と改善の方向性

キーワード検索

検索キーワード

辛いラーメンのお店

単純なキーワード検索は
表記ゆれを取りこぼす
字面で検索するから

- ✓ 辛いラーメンのお店
- ✕ 激辛ラーメンのお店
- ✕ 甘辛ラーメンのお店
- ✕ ただのラーメン店
- ✕ 洋菓子のお店
- ✕ 和菓子のお店

キーワード検索の課題と改善の方向性

セマンティック
サーチ

検索キーワード

辛いラーメンのお店

意味で検索するので
より広い範囲のデータを
取得できる。

- ✓ 辛いラーメンのお店
- ✓ 激辛ラーメンのお店
- ⓪ 甘辛ラーメンのお店
- ⓪ ただのラーメン店
- ✕ 洋菓子のお店
- ✕ 和菓子のお店

セマンティックサーチを 実現するために



- セマンティックサーチ
 - 検索クエリーの意味や文脈を理解して関連性の高い情報を提供する
 - キーワードの一致を確認する従来の検索とは異なる
- 意味や文脈の理解に必要な技術
 - テキストをベクトルデータへ変換：LLM（大規模言語モデル）
 - ベクトルデータとして扱うことで意味的に類似した文や単語がベクトル空間上で近い位置にマッピングされ、意味による比較が可能
- ベクトルデータの類似度を計算する技術
 - 検索時は検索対象のテキストと検索クエリーをベクトル化し類似度を比較するため
 - 膨大なベクトルの中から、類似するベクトルを探し出す -> 効率的なインデックス構造やアルゴリズムが重要

RaBitQを採用した理由

- RaBitQとは
 - 各32ビット浮動小数点数を0/1の1ビットで表現する
 - 単純計算でデータ量は1/32になる
 - ベクトルの量子化と呼ばれる技術の一つ
 - 計算対象が1ビットのデータになる
 - サイズの減少=メモリに載るくらいデータを小さくできる
 - ビット演算、SIMDを活用して検索処理を高速化している
- RaBitQを採用した理由
 - データ量が少ない！
 - 高速に計算できる！

PostgreSQLのセマンティックサーチ

- PostgreSQLセマンティックサーチを使うには？
 - pgvectorを使う
 - PGroongaを使う
- pgvectorとは？
 - 類似したベクトルデータを検索する機能を提供するPostgreSQLの拡張機能
 - 類似度の計算に様々な計算方法が多数提供されている（ユークリッド距離、マンハッタン距離、内積、コサイン距離、ハミング距離、ジャカード距離）
 - インデックスを使用して検索を高速化できる
- PGroongaとは？
 - 読み方：ピージーるんが
 - 全言語対応の超高速全文検索機能を提供するPostgreSQLの拡張機能
 - PostgreSQLのインデックスとして使える
 - PostgreSQLのデータを使って検索する = ゼロETLで利用できる
 - PostgreSQLのSQL構文をほぼそのまま使える = 学習コストが低い

pgvectorのベクトルデータ挿入

pgvectorの
ベクトルデータ挿入

テキストデータとベクトルデータを
一緒にINSERT

データ投入

辛いラーメンのお店
激辛ラーメンのお店
甘辛ラーメンのお店
ただのラーメン店
洋菓子のお店
和菓子のお店

ベクトル化

[0.112311..., 0.21088...]
[0.121311..., 0.21458...]
[0.125612..., 0.21333...]
[0.175212..., 0.20137...]
[0.375212..., 0.00137...]
[0.355012..., 0.01147...]

テキストデータ	ベクトルデータ
辛いラーメンのお店	[0.112311..., 0.21088...]
激辛ラーメンのお店	[0.121311..., 0.21458...]
甘辛ラーメンのお店	[0.125612..., 0.21333...]
ただのラーメン店	[0.175212..., 0.20137...]
洋菓子のお店	[0.375212..., 0.00137...]
和菓子のお店	[0.355012..., 0.01147...]

PGroongaのベクトルデータ挿入

PGroongaの
ベクトルデータ挿入

データ投入

辛いラーメンのお店
激辛ラーメンのお店
甘辛ラーメンのお店
ただのラーメン店
洋菓子のお店
和菓子のお店

テキストデータのみINSERT
= 普通にINSERTするだけ

テキストデータ	ベクトルデータ
辛いラーメンのお店	[0.112311..., 0.21088...]
激辛ラーメンのお店	[0.121311..., 0.21458...]
甘辛ラーメンのお店	[0.125612..., 0.21333...]
ただのラーメン店	[0.175212..., 0.20137...]
洋菓子のお店	[0.375212..., 0.00137...]
和菓子のお店	[0.355012..., 0.01147...]

ベクトルデータは
INSERT時に自動生成

性能

- 以下の内容について発表予定
 - INSERT速度の変化
 - インデックスサイズ
 - キーワード検索との検索結果の比較

用例：FAQ検索・ ナレッジベース検索の高度化

- 以下のような具体例を紹介予定
 - FAQ検索の例
 - よくある質問、似たような質問の検索
 - これらはキーワードベースではヒットしにくい
 - 質問内容は人により表現が様々なため
 - RAGとしての使用例