

Presearph.js

プライバシーを妥協しない推薦エンジンの開発

学習駆動コース社会実装ゼミ 宮沢 純正

1. TL;DR ~ これってなに? ~

プライバシーを守りつつ
「あなたへのおすすめ」の利便性も享受したい!
 ↑ これを実現したい ↑

2. 現状

- DuckDuckGoやStartPageなど、トラッキングを行わないことを特徴とした検索・サービスは登場しているが、まだまだ普及しているとは言えない印象
- 大規模サービスはユーザーの行動を追跡し、データを元にパーソナライズを実現しているが、プライバシーの懸念も高まっている。

3. 問題意識

プライバシーとパーソナライゼーションは
トレードオフ?

- プライバシー重視のサービスは利便性（≒パーソナライズ）に欠けることが多い
- トラッキングしないサービスへの移行はあまり進まない

4. 現時点のシステム構成

サーバーサイド

- 検索対象となる大量のページを保持
- 任意の指標（例：現在閲覧中ページとの類似度など）を用い、一度大まかに絞り込み
- タイトル・サマリーなどをクライアント側へ返す

クライアントサイド（ユーザー端末）

- 閲覧ページのEmbedding（文章ベクトル表現）を作成し、自分だけが持つプロフィールを更新
- サーバーが返した候補を「自分のEmbedding履歴」との類似度でリランキングする
- →これにより、サーバー側には個人の嗜好や詳細な行動データを渡さないが、ユーザーごとに最適化された「おすすめ」が可能になる

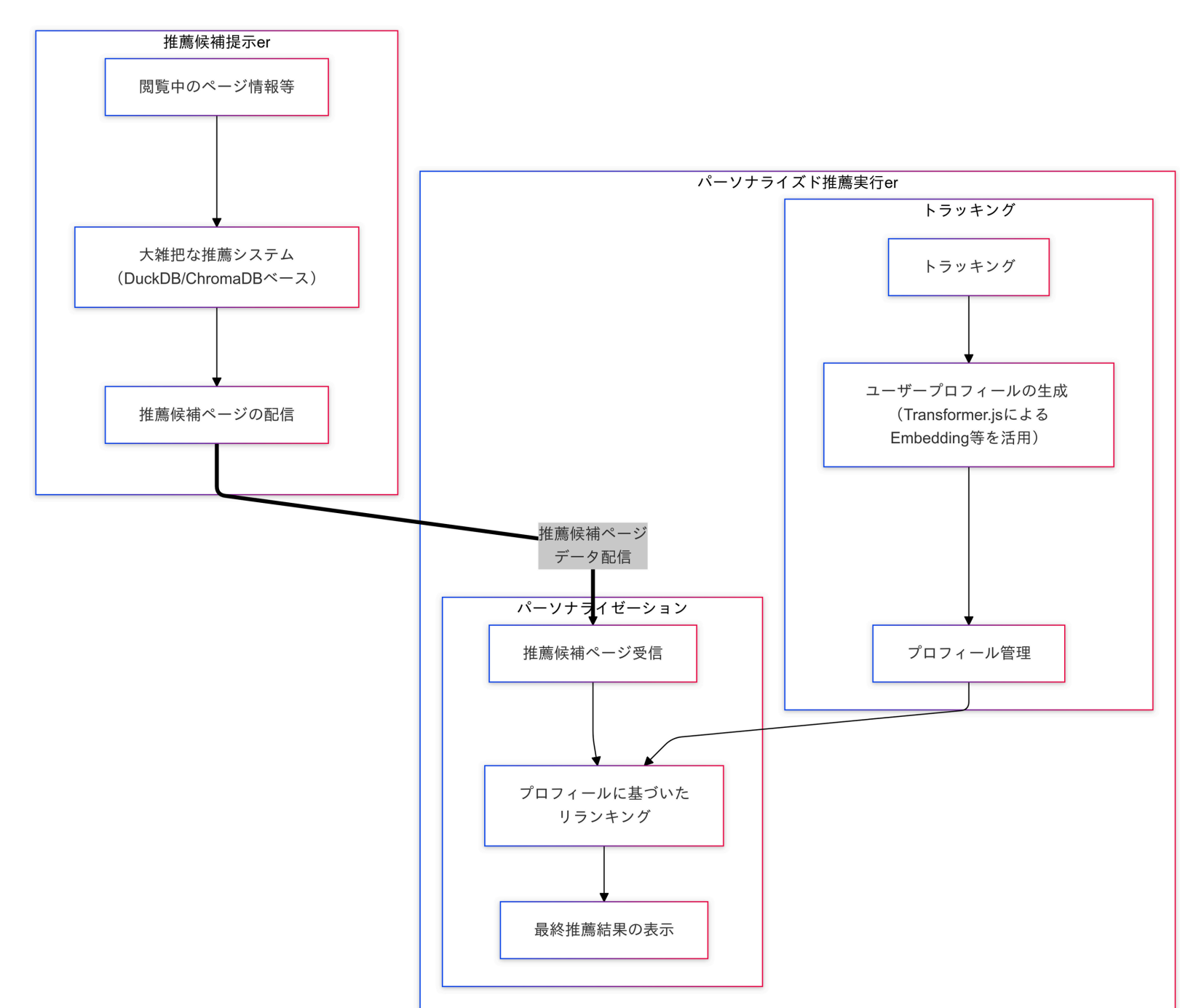


図1. 構成図

5. 現時点の試作・実装

1. WordPress向け プラグイン実装

- 「WordPressでパーソナライズされたおすすめ記事」を簡単に実現するプラグインを開発
 - WordPressは依然として国内シェアが高く、導入ハードルが低い
 - 公式プラグインディレクトリにも、現状はプライバシー保護+パーソナライズを両立する類似プラグインが存在しない

2. デモサイト「LibMeet」

- 書籍ジャンル「オライリー本」とユーザーの嗜好をマッチングさせるアプリ
- 本番環境での導入検証として、ユーザーのいいね=閲覧履歴として解釈した実装を実施

6. 今後の展望

- 推薦アルゴリズムの性能向上
 - Embeddingによる類似度検索に加え、連合学習を用いた協調フィルタリングの導入による精度の向上
- 導入フローをさらに簡便化
 - 現時点ではWordPressプラグインなどピンポイントの実装例がメイン
 - 汎用ライブラリとして切り出し、ドキュメントを整備・“10行程度であなたへのおすすめを導入”できる世界を目指す
- プライバシーモデルの明確化・保証
 - 完全にクライアント主導でデータを保持するモデルの安全性検証
 - ユーザーの知らないうちに情報が送信されない仕組みの透明化