

1. ビット列⇔URL対応辞書が必須

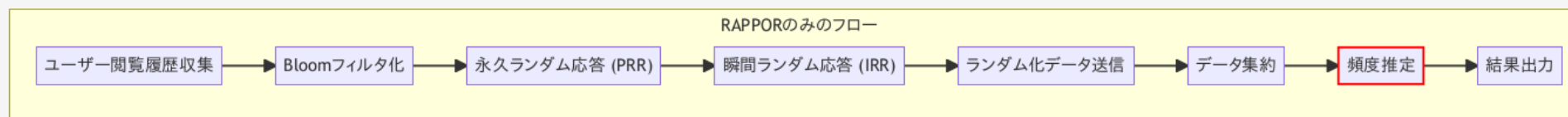
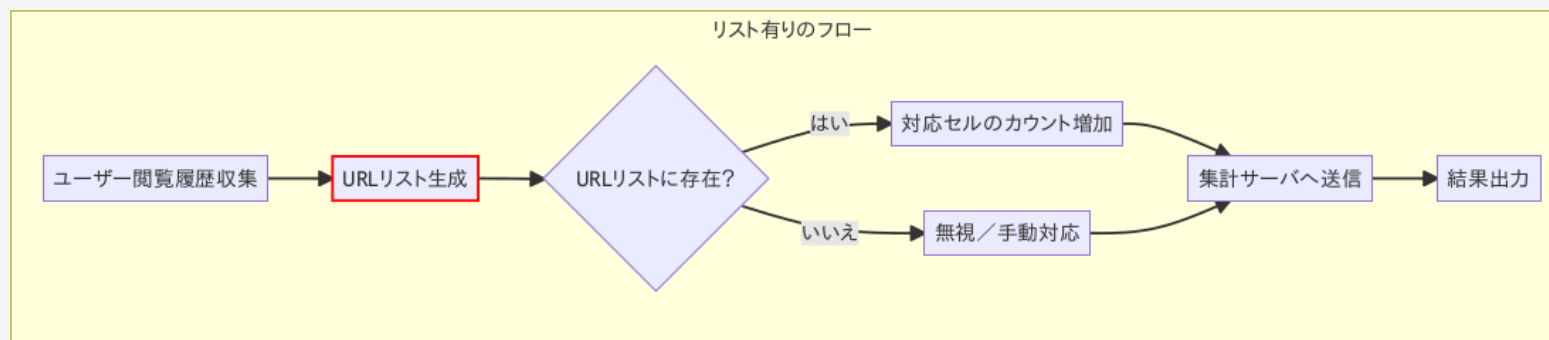
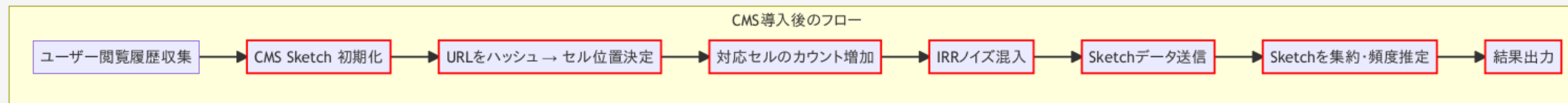
- RAPPORはBloomフィルタでURLをビット列に変換＋ランダム応答を行う
- どのビット位置がどのURLに対応するかを事前に知る辞書（URLリスト）がないと、ビット列からURLを特定・集計できない

課題: 無数のURLをどう扱うか

- ウェブ上には無数のURLが存在
- 事前に全URLをリスト化するのは不可能
- → 頻度集計のための候補リストが使えない

解決策の概要

1. 行×列の小さな表（CMS Sketch）を用意
2. URLごとに対応する「ひとつのセル」だけを+1カウント
3. 送信前にランダムノイズ（IRR）を混入してローカル匿名化
4. 集約後、Sketchから任意URLの訪問頻度を推定可能



```
flowchart TD
  subgraph CMS導入前
    A[ユーザー閲覧履歴収集] --> B[URLリスト生成]
    B --> C{URLはリストに存在?}
    C -->|はい| D[対応セルのカウント増加]
    C -->|いいえ| E[無視 or リスト追加のため手動対応]
    D --> F[集計サーバへ送信]
    E --> F
    F --> G[全URLリストを走査して集計]
    G --> H[結果出力]
  end

  subgraph CMS導入後
    A2[ユーザー閲覧履歴収集] --> B2[CMS Sketch 初期化]
    B2 --> C2[URLをハッシュ → 行×列のセル位置決定]
    C2 --> D2[対応セルのカウント増加]
    D2 --> E2[IRRによるノイズ混入]
    E2 --> F2[Sketchデータを送信]
    F2 --> G2[SketchをAggregate]
    G2 --> H2[estimate_url() で頻度推定]
    H2 --> I2[結果出力]
  end

  `` -->
```