

Wikipediaカテゴリを用いたブログ 著者の得意分野プロファイリング

東京大学大学院学際情報学府修士1年

野田陽平

東京大学情報基盤センター図書館電子化研究部門

清田陽司

中川裕志

本研究の目的

- Wikipediaのカテゴリやタイトルの関係を利用して、膨大なブログ投稿者の中から特定の分野に精通し、有用なブログ記事を投稿するブロガーを発見する。

関連研究

=既存のブログ記事分類=

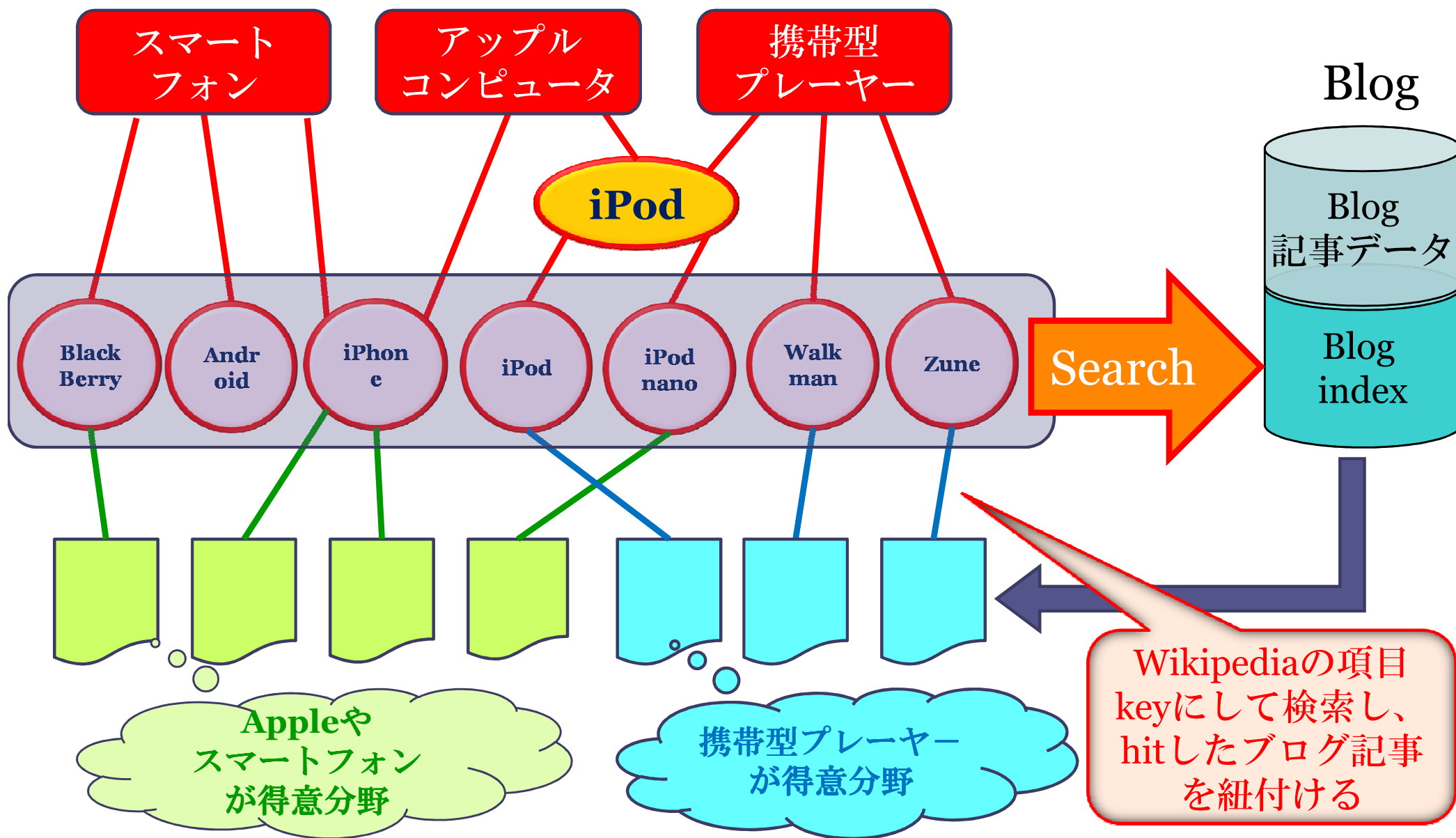
辞書を作成し、文書ごとに作成した単語ベクトルをもとに分類

=既存の α ブロガー発見手法=

リンク構造や特定の話題への言及のタイミングによる分析

本研究の手法

Wikipedia



まとめ

- Wikipediaのグラフ構造を使って、各分野に注力して投稿しているブログ投稿者を抽出することができた

今後の課題

- 時系列での各項目の頻度分析を使用
(特定の項目をコンスタントに記述しているブロガーなど)
- スプログへの対応
- より深いカテゴリ間情報の分析