

# 第1章 関連研究

## 概要

本章では、本研究に関連する研究領域について整理し、本研究の特徴や位置づけについて述べる。

## 1.1. 情報検索

情報検索のアルゴリズムについては、文献 [?]、[?] で詳しく解説されている。また、検索結果を改善するための研究でよく用いられる手法としては、サポートベクターマシン [?] などの機械学習の手法がよく用いられる。

検索アルゴリズムを提案することによって、検索システムの性能や精度を向上させようとしている研究は多い。

また、ソーシャルな仕組みを利用したりなど人力を利用した検索システムの提案も多い。

莊司らは、発見した Web ページを読んだ際に、読み手がどのように感じるかをクエリとして入力可能な Web 情報検索システムを提案している [?]

村田らは、検索エンジンのクリックログと検索結果として表示されるページのタイトル、スニペットを利用して、検索結果ランキングを生成するシステムを提案している [?]

松井らは、情報と人を瞬時に発見することによって、検索サービスとソーシャルサービスの双方の利点を活用できる検索システムを提案している。[?]。このシステムを利用することにより、ユーザは、検索結果と同時にページの閲覧者も発見できる。さらに、閲覧者とチャットによるコミュニケーションを行うこともできる。

White らは、ユーザのアクセス履歴を利用して、ページの重要度を決定し、人気の高いページをランキングする手法を提案している [?]

Agichtein らは、大規模な実験により、ユーザの振る舞いの情報を利用することによって、Web 検索の結果を改善することが可能であることを示している [?]

Morris らは、検索キーワードを共有したりなど、数人の小規模なユーザ間での協調作業により、情報検索の効率向上を目指している [?]

さらに、新しい検索インタフェースを提案することによって、ユーザにとってやさしい検索システムを提案しようという研究事例もある。

吉田らは、検索結果ページ中に現れる重要語を話題語として抽出し、マウスによる単語のドラッグ&ドロップによって、AND 検索、OR 検索、NOT 検索を直感的に行い、検索結果の再ランキングをできる手法を提案している [?]

## 1.2. 情報フィルタリング

情報フィルタリングの手法を分類すると、主にコンテンツに基づくフィルタリングと協調フィルタリングに分類できる [?]

コンテンツに基づくフィルタリングとは、コンテンツのコンテンツに基づき情報の取捨選択を行う手法のことである。コンテンツに基づくフィルタリングに関する研究の例として、文献 [?], [?] があげられる。

協調フィルタリングとは、ユーザの過去の行動を記録し、そのユーザと類似した行動をとっているユーザの嗜好情報から、ユーザに情報を推薦する手法のことである。協調フィルタリングに関する論文の例としては、文献 [?], [?], [?] があげられる。また、協調フィルタリングの実用例としてもっとも有名なのが Amazon の推薦システムである [?]

また、近年では、コンテンツに基づくフィルタリングと協調フィルタリングを組み合わせたハイブリッド法 [?] に関する研究もさかんになってきている [?]

### 1.3. ソーシャルブックマークを利用した研究

ソーシャルブックマークデータを利用した関連研究について紹介する．ソーシャルブックマークデータを利用した先行研究としては，主に以下の4種類の研究に分類できる．

1. Web ページの検索
2. Web ページの推薦
3. ソーシャルブックマークユーザの推薦
4. ソーシャルブックマーク，Folksonomy の分析

上述の中で本研究は，Web ページの検索に分類されるが，Web ページの推薦という側面も持っている．

#### 1.3.1. Web ページの検索

ソーシャルブックマークデータを用いた Web ページの検索に関する関連研究としては，以下のような研究事例がある．

Xu らは，ソーシャルブックマークデータを利用したパーソナライズド検索手法を提案している [?]．さらに，ソーシャルブックマークのタグ情報を利用することによって，パーソナライズド検索を自動的に評価する手法も提案している．Yanbe らは，ソーシャルブックマークのブックマーク数を新たな指標 SBRank を提案し，PageRank と SBRank を統合して，Web 検索ランキング精度の向上を計っている [?][?]．Takahashi らは，ソーシャルブックマークデータの時間データを利用して，鮮度の高い Web ページを取得する検索手法を提案している [?]．ここでは，ブックマーク日時の散らばりの大きさから，Web ページの賞味期限を判定して，賞味期限を過ぎていない鮮度の高い Web ページを取得している．この研究は，ソーシャルブックマークの時間データに着目した情報検索手法を提案しているため，本研究と最も近い．本研究との違いは，この研究では，鮮度の高い Web ページを取得することを目的としているのに対して，本件研究では，体系だった知識をえられ，長期に渡って役立つ Web ページを取得することを目的としている点である．そのため，時間データの利用の仕方も異なっている．

#### 1.3.2. 情報推薦

情報推薦に関する研究としては，以下のような研究事例がある．

Niwa らは，タグのクラスタリングをおこなうことによりタグの表記ゆれの問題の解決をはかり，ユーザのブックマーク情報からユーザの趣向に沿った Web ページの推薦をおこなう手法を提案している [?]．Sasaki らは，タグを表象とする Web コンテンツ群の類似性に基づいた Web コンテンツ推薦システムを提案している [?]．

#### 1.3.3. ソーシャルブックマークユーザの推薦

ソーシャルブックマークユーザの推薦に関する研究としては，以下のような研究事例がある．

白土らは，ソーシャルブックマークユーザのブックマーク情報からユーザの関連度を解析した結果から興味の類似したユーザを推薦し，ネットワーク図として表示するシステムを構築した [?]．大力らは，ソーシャルブックマークユーザの中のイノベータ，いわば ブックマーカーに注目した情報推薦手法を提案している [?]．

#### 1.3.4. ソーシャルブックマーク, Folksonomy の分析

ソーシャルブックマークや Folksonomy の分析に関する研究としては、以下のような研究事例がある。

Golder らは、ソーシャルブックマークのユーザやタグ、ブックマークの性質について分析し、各 Web ページに対する各タグの出現頻度は一定値に収束することを証明している [?]。Paul らは、del.icio.us のデータを収集して、ソーシャルブックマークが Web 検索において大きな改革を起こせるかどうか検討している。その結果、現状ではソーシャルブックマークのデータ量不足の問題やタグのゆらぎの問題から、現時点のデータでは、Web 検索に関して劇的な改革は起こせないが、今後ソーシャルブックマークのデータ量が急激に増えたりした場合は、Web 検索において改革を起こせる可能性があると結論づけている [?]。川中らは、あるタグと共起関係の強いタグを取得し、出現時期の早いほうを親タグとする手法を用いることによって、タグの時系列の関係性をグラフ化している [?]。

#### 1.4. 検索キーワード

検索結果ではなく、検索クエリを改善することによって、Web からの検索を楽にしようという研究事例も多い。

クエリ拡張の手法として、Robertson らが提案した、検索結果から適当な単語を抽出して検索質問拡張を行う手法の一つである RSV(Robertson's Selection Value) がある [?]。

また、正田らは、ユーザが与えたクエリでの検索結果上位 R 件を適合文書、それ以下を不適合文書として、上述の RSV を用いてクエリ拡張を行い、新たなクエリの重みにより初期の検索結果をソートする手法を提案している [?]。

大塚らは、大規模なアクセスログから抽出されたユーザの Web 検索の検索単語と、それによって閲覧したページを解析することで、関連語を抽出する手法を提案している [?]。

大石らは、ユーザの意図する検索クエリを生成するための方法として、センテンス間の距離に注目した関連単語抽出アルゴリズムを提案する。このアルゴリズムは重要な語の近くに出現する単語は重要であるという考えに基づいている [?]。

今井らは、入力した検索クエリが多義語の場合、選択された URL に基いて推薦クエリを生成する手法を提案している [?]。

木田らは、時間とともに変化するクエリ間類似度を利用して、検索クエリをクラスタリングする手法を提案している [?]。

Zhang らは、検索エンジンのキーワード補完のように関連するクエリを推薦するために、クエリの分類と関連クエリを発見するための手法を提案している [?]。この研究では、キーワード間の TFIDF 類似度の代わりに、SF(Search Frequency) を用いた SFIDF を基づく Content Similarity の線形結合を類似度としたクラスタリング手法を提案している。

Cao らは、ユーザーの入力したクエリとクリックされた URL をクラスタリングしたものと同一セッション内で入力されたクエリの関連性に基づいてクエリ推薦を行う手法を提案している [?]。

甲谷らは、検索クエリとクリックされた URL 情報を利用して、Web サイトに到達するために頻繁に使用されるクエリを発見することで、クエリ推薦を行う手法を提案している [?]。

安川らは、検索クエリのログから検索語の関連語を取得して、関連語のみに限定した単語方向のクラスタを生成し、クラスタと Web ページ群との対応をユーザに提示する Web 検索の手法を提案している [?]。

## 1.5. ファイルアクセスに関する研究

デスクトップ上のファイルアクセスを容易にするための研究も多くなされている。Rekimoto らは、PC 上のファイルは、すべてデスクトップに置かれ、時間とともに消えていくシステムを提案している [?]. このシステムでファイルを探す場合は、過去のデスクトップに遡っていく。そして、そのときデスクトップに置かれていたファイル群と一緒にファイルを見つけることができる。

Soules らは、複数のファイル同士の関連性を利用して、ファイル検索の精度をあげる手法を提案している [?].

## 1.6. 情報の管理に関する研究

## 1.7. ロングセラーに関する研究