平成 24 年度

フロンティアプロジェクト

修士学位論文

ユーザの嗜好に適応した Web ページ推薦システムの研究

Research of system that recommends web page to suit one's taste

1155086　　駒木　快比古

指導教員　　清水　明宏

2013 年 2 月 12 日

高知工科大学 フロンティア工学コース

要 旨

ユーザの嗜好に適応した
Web ページ推薦システムの研究

駒 木　快 比 古

組織において，組織に属する人間が知識を深めていくために，個人が持つ知識・情報を組織全体で活用し，新たな知識・情報を生み出すことは非常に重要視されている．古くから組織の中での情報共有は行われており，その形態は様々である．近年では情報共有を IT 技術を用いてシステム化することによって，より効率的に情報を共有する手法が提案されている．

既存研究として，共有するために蓄積されたブックマークを，利用者の嗜好に合わせて推薦するシステムがある．既存システムの問題点として，蓄積されたブックマークが分類されておらず，利用者にとって有用なブックマークが推薦されづらい点があげられる．

本研究では，蓄積されたブックマークを自動分類し，既存システムと比べ，より利用者の嗜好に合わせたページを推薦できる検索支援システムを提案，開発した．提案システムを運用し，既存システムと比較評価した．その結果，既存システムより，利用者の嗜好に合わせた情報を推薦することが出来た．

キーワード　情報共有, 検索支援, ソーシャルブックマーク

Abstract

Research of system that recommends web page to suit one's taste

Yoshihiko Komaki

In an organization, it is very important that we make use of the information using personal knowledge in the organization in order to obtain knowledge. Since early times, various forms of information sharing is done. In recently, the method which efficiently shares information is proposed by information sharing system using IT tecnology.

In existing research, there is a system recommended shared bookmark to the hobbies and diversions of the user. As a problems, shared bookmark is not classified. So right recommendation is not performed.

In this thesis, developed a system classifies shared bookmark automatically, and developed and suggested a system more recommended shared bookmark to the hobbies and diversions of the user. We compared the suggestion system with the existing system. As a result, suggested system was better than an existing system.

key words Information sharing, retrieval support, Social Bookmark

目次

第 1 章　はじめに　1

1.1　背景と目的 1

1.2　本論文の概要 2

第 2 章　組織における情報共有　3

2.1　情報共有の必要性 3

2.2　共有する情報の種類 3

2.2.1　電子掲示板 4

2.2.2　ブログ 4

2.2.3　マイクロブログ 4

2.2.4　ブックマーク 5

2.3　ブックマークを用いた情報共有 5

第 3 章　既存方式　7

3.1　ブックマーク情報を用いた Web 検索支援システム 7

3.2　システム詳細 8

3.2.1　ブックマークの収集と分析 8

3.2.2　利用者の嗜好分析 9

3.2.3　利用者のグループ化 9

3.2.4　ブックマークの推薦 9

3.3　既存方式の評価と問題点 10

3.3.1　評価 10

3.3.2　問題点 10

第 4 章　提案方式　12

4.1 システム概要 12
4.2 ブックマークの収集と分析，嗜好タグの付与，嗜好の似通った利用者同士のグループ化 13
4.3 ブックマークされたページの内容に沿った分類 13
4.3.1 単語抽出 14
4.3.2 表記揺れの対応 14
4.3.3 無意味な単語の除外 14
4.3.4 クラスタリングによる分類 15
4.3.5 分類対象 16
4.4 ブックマークの推薦 18

第 5 章 評価と考察 **20**
5.1 分類の精度 20
5.2 利用履歴からの評価 20
5.3 考察 21

第 6 章 おわりに **22**

謝辞 **23**

参考文献 **24**

図目次

2.1 情報共有システム概要図 4
2.2 ソーシャルブックマーク概要 6

3.1 既存システム概要 7
3.2 既存システムの実行画面 10

4.1 提案システム概要 13
4.2 グループが持つブックマークの分類 17
4.3 既存システムの実行画面 19

表目次

3.1 ブックマーク情報 8
3.2 利用者の嗜好 9

4.1 「PHP」タグが付与されたブックマークの分類例 17

第 1 章

はじめに

本章では，本研究における社会的な背景と目的について述べ，本論文の概要について述べる．

1.1 背景と目的

近年，研究室や会社の部署などの組織が行う情報収集の活動において，新聞や書籍・雑誌などの物理媒体からだけでなく，ニュースサイトやブログ，掲示板などの Web 上からの情報収集は欠かせないものである [1]．また，情報通信技術の普及に伴い，Web 上の情報は急速に増加している．また，世界中のドメイン数は 2012 年時点で約 9 億個存在し，Web ページ数は 2008 年時点で 1 兆ページ存在しており，今も 2 次関数的増加をしている [2][3].

この膨大な情報の中から，組織で行われる情報収集の手段として，キーワードを用いた Web 検索や，RSS リーダを用いた情報収集が挙げられる．しかし，キーワードを用いた Web 検索では，Web 上の情報量が多すぎるため，情報の選別に時間がかかってしまう．RSS リーダは人手で作成している Web ページなどでは，そのシステム上コストがかかるためあまり使われていない．また RSS により得られたページは，登録しておいたサイトが更新する度に一方的に与えられるページであり，整理されておらず，過去の情報を取得し辛いという問題点がある [4].

そこで組織では，お互いが持つ情報を共有し，その情報から自分に有益な情報を選び出す情報共有システムが使われている．情報共有システムの 1 つに，ブックマークを利用した共有システムがあげられる [5]．ブックマークを利用した共有システムを利用する際に，自身

に適したものを選び出し，新たな知識として学習するためには，登録されたブックマークを網羅的に閲覧しなければならない．

本研究では，組織内でブックマークを用いた情報共有の中から有用な情報を選び出す際に，自身に適した情報の検索支援を行うシステムを提案，開発する．

1.2　本論文の概要

本論文では，組織内でブックマークを用いた情報共有の中から有用な情報を選び出す際に，より自身に適した情報の検索支援を行うシステムについて述べる．

第二章では，組織における情報共有の重要性と，ブックマークを用いた情報共有について述べる．

第三章では，ブックマークを用いた情報共有システムに対する既存システムとその問題点について述べる．

第四章では，既存システムを応用したブックマークを用いた情報共有システムを提案し，その具体的なシステム構成についても述べる．

第五章では，提案システムの運用評価と，評価結果からの考察を述べる．

最後に，本論文のまとめと今後の課題について述べる．

第 2 章

組織における情報共有

本章では，まず一般的に組織で行われている情報共有について述べる．そして，ブックマークを用いた情報共有システムについて説明する．

2.1　情報共有の必要性

組織において，組織に属する人間が知識を深めていくために，個人が持つ知識・情報を組織全体で活用し，新たな知識・情報を生み出すことはとても重要視されている．古くから組織の中での情報共有は行われており，その形態は様々である．ある会社の部署において，業務の中で同じ作業を行うチームを編成し，お互いにノウハウを教え合うなど，直接的なコミュニケーションを通したものや，引き継ぎ資料などを用いた間接的なコミュニケーションによって情報共有を行う手法などがある．このように組織内で情報共有することで，個人の活動範囲では知りえなかったことや出来なかったことを，他の人間が持つ知識・情報により成し遂げ新たな知識・情報とする取り組みがなされている．

近年では情報共有を IT 技術によってシステム化することによって，より効率的に情報を共有する手法が提案されている．IT 技術による情報共有システムの概要を図 2.1 に示す．

2.2　共有する情報の種類

また，IT 技術の発達に伴い，情報共有にも様々な手法が存在するようになってきた．その中でも代表的なものを以下に 3 つ述べる．

図 2.1　情報共有システム概要図

2.2.1　電子掲示板

複数の利用者が同じ場所に情報を追加することで情報を共有する手法である．情報入力は，入力フォームを通して行われる．閲覧時には，データベースに蓄積された情報を加工せず，入力された順に情報を表示する．電子掲示板の特徴として，一つの話題について複数人が集まり議論を行うことで，それぞれの知識を引き出し合い，他では得ることの出来ない情報が得られる可能性がある．その反面，情報が発散しやすいという欠点がある．

2.2.2　ブログ

ブログはウェブログの略であり，個人が自由に情報を入力する手法である．時系列にページを自動生成する機能や，他のブログ記事との連携機能，コメント機能を有している．自身が持つ知識・情報を，ブログ記事として作成し，今までその知識・情報を知りえなかった人に対して情報を提供することが出来る．ブログは個人が自由に情報を入力出来る半面，自身が持つ知識を言語化しなければならず，それ自体が煩雑な作業であり，また知識を言語化することに慣れていない人にとってはブログ記事を書くこと自体が難しい．

2.2.3　マイクロブログ

ミニブログとも呼ばれるものであり，ブログに比べて文章量が短く手短に情報を発信することが出来るのが特徴である．また，マイクロブログ利用者同士での交流も取れるように

なっている．マイクロブログもブログ同様，自身が持つ知識を言語化しなければならないことや，仕様上文章量に制限があり，一定の文章量の情報を発信することができない．これにより，詳細な内容を 1 つのページで発信することが出来ず，複数のページによって分割され後に情報を確認する際に面倒な手順を踏まなければならなくなる．

2.2.4　ブックマーク

ブックマークとは，Web 上にある膨大な情報の中からページを探し出し，そのページを探し出した人自身が選別，閲覧し，有用であると判断したページを，ブラウザに保存する機能，またはそのページのことである．このブックマークは，電子掲示板やブログと異なり，有用だと判断された知識・情報が既に言語化されており，且つある事柄について書かれていて，その内容が後から発散するということが無い．またブックマークとは普段閲覧しているページの中から有用だと判断されたページなので，利用者が時間をかけて別のことをするということが無く，気軽に登録することが出来る．またブックマークは，ユーザの趣味嗜好により変わってくるものであり，そのユーザの趣味嗜好を指し示す指標として利用することも出来る．

2.3　ブックマークを用いた情報共有

近年，組織ではブックマークを用いた情報共有システムが使われている．組織とは，知識活動の背景が似通っている人物の集合体，また共通の目標を有し，目標達成のために協働する人物の集合体である．そのため，組織内の 1 人がブックマークしたページ，つまり有用であると判断されたページは，同組織に属する，ブックマークをしたユーザ以外の人間にとっても有用であると判断される可能性が高い．また，ブックマークとは，その人間が新たに得た情報であるため，同組織内の共に活動する人間が，今何に興味を持っているのか，どのような情報を得たのか，ということを知ることが出来る．

組織でブックマークを共有する手法として，ソーシャルブックマークがあげられる．ソー

2.3　ブックマークを用いた情報共有

シャルブックマークとは，個々の利用者が自分のブックマークを Web 上に保存し，他の利用者に公開することでブックマークの共有を可能にしたサービスである．日本国内では，多くのソーシャルブックマークが展開されている．中でも Livedoor クリップや，はてなブックマークが有名である．利用者は，自身のブックマークにタグと呼ばれる分類用の語句を付与し，ソーシャルブックマークに登録する．タグは，ブックマークの特徴を表す単語である．ソーシャルブックマークの概要図を図 2.2 に示す．

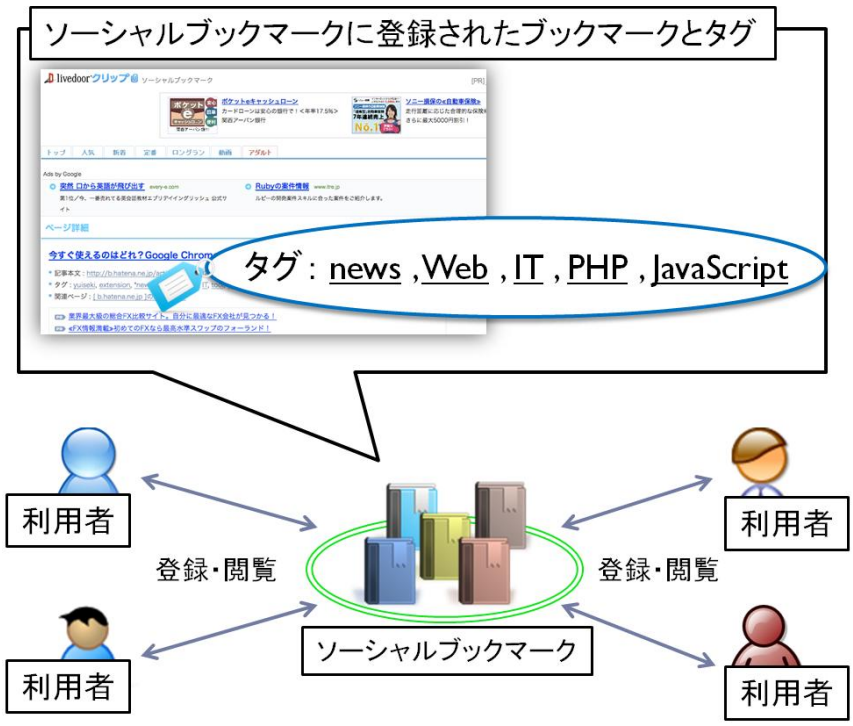

図 2.2　ソーシャルブックマーク概要

利用者は，ソーシャルブックマークに登録されたブックマークの中から，自身に有用だと判断したものを抜き出して，自身の知識・情報とする．しかし，ソーシャルブックマークに登録されたブックマークが増えると，自身に有用なブックマークを選び出すこと自体が煩雑な作業となってしまう．

第 3 章

既存方式

本章では，組織内でブックマーク共有システムを用いたときの，利用者の嗜好に合わせた検索支援システムについて，森らの方式について述べる [6].

3.1　ブックマーク情報を用いた Web 検索支援システム

既存方式では，ソーシャルブックマークに登録された各利用者のブックマークを推薦する．これにより一度は他人に評価された情報を推薦するため，より有益な情報推薦が可能となる．同時に Web 上の膨大な情報を絞り込むことが可能となる．既存システムの概要を図 3.1 に示す.

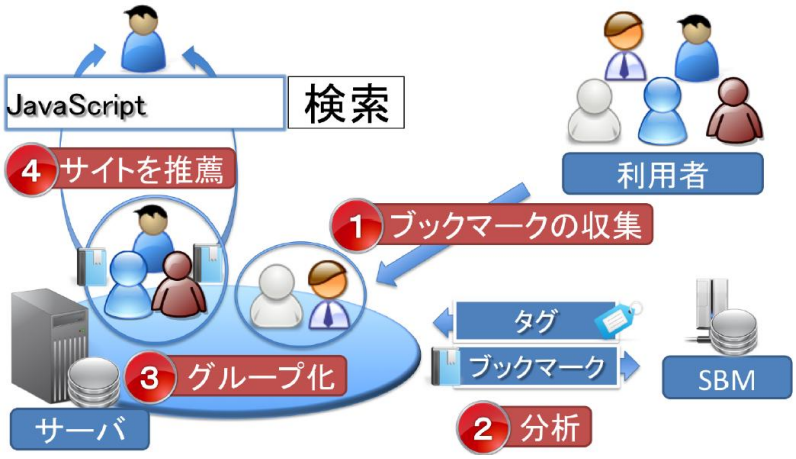

図 3.1　既存システム概要

既存方式は大きく分けて次の 4 つのステップで構成される．まず既存方式の利用者からブックマーク情報を収集する．次にソーシャルブックマークによりブックマーク情報を分析

する．次にブックマーク情報を元に各利用者の嗜好を分析し，嗜好の似通った利用者同士をグループ化する．最後に利用者が検索をおこなう際に，同じグループに所属する他の利用者のブックマークを推薦する．

システムの詳細を各ステップごとに説明する．

3.2　システム詳細

ここでは，既存方式のシステムで行われている，ブックマークの収集と分析，利用者の嗜好分析，利用者のグループ化，ブックマークの推薦について詳細を述べる．

3.2.1　ブックマークの収集と分析

利用者のブラウザ上のブックマーク情報を収集する．そしてソーシャルブックマークにアクセスし，タグの情報を抽出する．ブックマークの情報は，表 3.1 のように，そのブックマークを保持している利用者 ID と URL とブックマークに付与されているタグから構成される．これらを一つのデータとして既存システムのサーバに保存する．

表 3.1　ブックマーク情報

利用者 ID	ブックマークの URL	タグ
A	http://www.xxxx...	PHP 開発 Java オブジェクト指向 ...
A	http://www.xxyy...	Java 開発 継承 eclipse SQL ...
B	http://www.xyzx...	イタリア 旅行 ツアー ヨーロッパ ...
C	http://www.zzzz...	Apple Mac iPod 音楽 iTunes ...
...	...	...

3.2.2　利用者の嗜好分析

まず，分析する利用者 ID のブックマーク情報を全て抽出する．そしてブックマークに付与されている各タグの出現回数を分析する．出現回数の多い上位 30 個のタグを利用者の嗜好と定義する．既存システムでは，表 3.2 のように利用者 ID とその利用者の嗜好タグをサーバ上に保存する．

表 3.2　利用者の嗜好

利用者 ID	嗜好タグ
A	PHP 開発 JavaScript プログラム SQL MySQL Apache Java ...
B	旅行 ツアー ヨーロッパ イタリア イギリス 絵画 イタリア 料理 ...
C	Java オブジェクト指向 eclipse 開発 リファレンス ...
...	...

3.2.3　利用者のグループ化

グループ化は，各利用者の嗜好 30 タグを比較し，14 個以上一致する利用者同士を同じグループとする．既存システムでは，14 個以上一致する利用者同士をグループ化したとき最も良い結果が得られている．

3.2.4　ブックマークの推薦

既存システムを用いて検索をする際の，ブックマーク推薦について説明する．まず，検索者は欲しいと思った情報についてのキーワードを入力する．次に検索者と同じグループにいる利用者のブックマーク情報を参照する．ブックマーク情報に付与されているタグと検索されたキーワードを比較する．タグと一致した場合，そのタグが付与されているブックマークを検索者に推薦する．キーワードを入力し，ブックマークが推薦される様子を図 3.2 に示す．

図 3.2　既存システムの実行画面

3.3　既存方式の評価と問題点

ここでは既存方式の評価と問題点について述べる．

3.3.1　評価

評価は，7 名の被験者にシステムを利用してもらい，その利用履歴から評価している．それぞれの利用者に推薦されたブックマークの中から，30 件のブックマークを閲覧し，有用だと判断したページの数をカウントした．そして，以下の評価式に代入して，有用であるページの割合を算出した．評価式は式 (3.1) とした．

$$\text{有用率} = \frac{\text{有用だと判断したブックマークの数}}{30} * 100 \tag{3.1}$$

その結果，7 名の被験者によって算出された有用率の平均が 47.6%となった．

3.3.2　問題点

評価の有用率より，利用者に推薦されたブックマークの約半数のブックマークが有用ではないと判断されたページであることが分かる．有用なページが推薦されなかった原因として，タグによる人物のグループ化のみで，ブックマークされたページの内容に沿っての分類が行われなかったため，必要の無い情報まで推薦されたと考えられる．

3.3 既存方式の評価と問題点

例えば，普段から PHP のソースコードばかり見ているユーザに対して，PHP という言語の概要や，使い方の説明が記載されたページを推薦しても有用だと判断されない．

第 4 章

提案方式

本章では，既存方式により推薦されるブックマークを更に分類し，不要な情報を取り除き，よりユーザの嗜好に合ったブックマークを推薦するシステムを提案する．

4.1 システム概要

既存方式により，ブックマークに付与されたタグを元に，嗜好の似通った利用者同士のグループ化がされている．推薦されたブックマークは，そのグループに属する人物が持つブックマークを推薦したもので，ブックマークされたページの内容に沿っての分類がされていなかった．提案システムでは，グループ化された人物が持つブックマーク全てを，その内容に沿って分類し，既存方式では不要とされていたブックマークを取り除く．そして，利用者の嗜好にあったブックマークを推薦することで，より利用者にとって有用なブックマークを推薦することが出来る．提案システムの概要を図 4.1 に示す．

提案システムは大きく分けて 3 つのステップで構成される．まず既存システムにより，利用者のブックマークの収集と分析，嗜好タグの付与，嗜好の似通った利用者同士のグループ化を行う．次に，グループに属する利用者がブックマークしているページを，ページの内容に沿って分類する．最後に，分類したブックマークの中から，利用者の嗜好に合うものを選び出し，ブックマークを推薦する．

今回はブックマーク情報として Livedoor が提供する研究用データを使用した [7]．システムの詳細を各ステップごとに説明する．

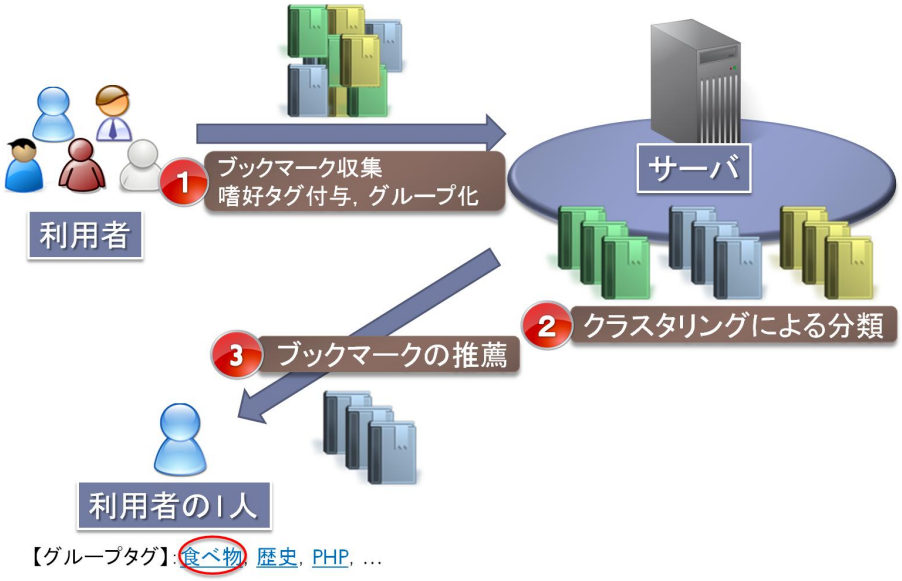

図 4.1　提案システム概要

4.2　ブックマークの収集と分析，嗜好タグの付与，嗜好の似通った利用者同士のグループ化

ブックマークの収集と分析，嗜好タグの付与，嗜好の似通った利用者同士のグループ化については，既存方式でされているので，3.2.1 節から 3.2.3 節を参照して頂きたい．また，このときグループ化する際に，他のユーザと一致した嗜好タグをグループタグとしてグループに付与し保存しておく．グループタグは，グループに属するユーザに付与された嗜好タグの一部でもあるので，それぞれのユーザはグループタグについて記載されたブックマークを多数所有していることになる．

4.3　ブックマークされたページの内容に沿った分類

既存方式ではタグを用いて人物のグループ化を行っていた．しかし，ブックマークに付与されているタグだけでは，ブックマークされたページの内容まで詳しく知ることが出来ず，ブックマークを細かく分類することが出来ない．そのため，提案システムでは，ブックマークされたページの本文を分析して分類を行う．

4.3.1　単語抽出

まずブックマークのソースを抽出し，タグ情報や文書型宣言を除く，そのページを意味する意味のある文章に対して形態素解析し，単語を抽出する．形態素解析には，日本語自然言語処理システムである「茶筌」を用いた [8].

既存システムによりグループ化された，グループに属する利用者のブックマークに対して分類を行う.

4.3.2　表記揺れの対応

言葉には様々な表現の種類が存在する．例えば,「日本」を意味する単語であれば,「にほん」「Japan」「JPN」「やまと」など複数存在する．このように同じ意味を持つ単語，同意語や同義語を別の種類で表現されることを表記揺れと言う．後に説明する分類手法において，この表記揺れに対応せずに処理を行った場合，正しくページを分類することが出来なかった．分類による精度を上げるため，この表記揺れに対応するためにシソーラスを用いた.

シソーラスとは同意語や同義語を同一の単語として扱うことの出来る辞書である．これにより，同意語や同意語を別の単語として処理されることが無くなり，分類による精度を上げる.

4.3.3　無意味な単語の除外

助詞や助動詞など，直接文章の意味を成さない単語や，Web ページのコメント欄に存在する世間話やスパム，分類すべき全ての対象に存在する単語（PHP に関するページを分類する際の「PHP」という単語）はページを表すものとしての意味を成さない．これらの単語を除外するために，tf-idf 法を用いる.

tf-idf 法とは，そのページを表す重要単語の抽出手法として利用されている [10]．ある文書中の単語の出現頻度を tf(term frequency)，全文書中における単語が出現する文書の割合を idf(inverse document frenqency) とし，両者の積がその文書における単語の重要語とな

る．それぞれ tf の算出は式 4.2，idf は式 4.3 を用いる．文書 D における単語 T の出現頻度 $tfreq(T,D)$，単語 T を含む文書数を $dfreq(T)$，全文書数を M とすると，単語 T の文書 D における重要度 $w(T,D)$ は式 4.1 のように定義される．

$$w(T,D) = tf(D,T) \cdot idf(T) \tag{4.1}$$

$$tf(T,D) = \frac{log(tfreq(T,D))}{log(tnum(D))} \tag{4.2}$$

$$itf(T) = log(\frac{M}{dfreq(T)}) \tag{4.3}$$

ここでは，無意味な単語の除外を行いたいので，この式 4.1 の $w(T,D)$ の重要度が，全ての重要度を降順に並べたときに下から 15 %に含まれる単語を無意味な単語として除外した．

4.3.4　クラスタリングによる分類

分類には，クラスタリング手法である Non negative Matrix Factorization（NMF）[9] を用いる．クラスタリングとは，与えられたデータを類似度に従って，クラスタと呼ばれる枠組みに自動的に分類する手法である．NMF は高次元ベクトル間の位置関係をできるだけ保存した形で，より低次元のベクトルへ変換する次元縮約を用いたクラスタリング手法である．文書クラスタリングにおいては，単語数が特徴を表す次元になる．そのため，非常に高次元なベクトルを処理することになり，精度や処理時間が膨大になる．NMF を使用することで，クラスタリングに影響の少ない特徴を削減し，精度の向上を処理時間の高速化が図れる．NMF は $m \times n$ の索引語文書行列 X を，$m \times k$ の行列 U と $n \times k$ の行列 V を転置した行列 VT の積に分解する．k はクラスタ数を表す．

$$\mathrm{X} = UV^T$$

NMF はクラスタに応じたトピックの次元を k 個と想定し，その基底ベクトルの線形和によって文書ベクトルと索引語ベクトルに対応する．基底ベクトルの係数がトピックとの関連度を表し，行列 V がクラスタリング結果となる．具体的には，i 番目の文書 d_i は，行列 X の第 i 列のベクトルで表現され，次元縮約された結果が，行列 V の第 i 行のベクトルとなる．このとき，V の第 i 行のベクトルは

$$(v_{i1}, v_{i2}, \cdots, v_{ik})$$

と表すことができる．与えられた索引語文書行列 X から，U と V は以下の繰り返しで得られる．

$$u_{ij} \leftarrow u_{ij} \frac{(XV)_{ij}}{(UV^TV)_{ij}} \tag{4.4}$$

$$u_{ij} \leftarrow u_{ij} \frac{(X^TV)_{ij}}{(VU^TU)_{ij}} \tag{4.5}$$

ここで u_{ij} と v_{ij} はそれぞれ U と V の i 行 j 列の要素を表している．また，$(X)_{ij}$ によって行列 X の i 行 j 列の要素も表している．繰り返しが終了した後に U を以下のように正規化する．

$$u_{ij} \leftarrow \frac{u_{ij}}{\sqrt{\sum_i u_{ij}^2}} \tag{4.6}$$

4.3.5　分類対象

あるグループに,「PHP」「歴史」「音楽」「食べ物」というグループタグが付与されているとき，それぞれのグループタグについて記載されたブックマークの分類を行う．初めに，このグループに属する利用者のブックマークの中から「PHP」タグが付いているブックマークを全て取り出す．そして，抽出した単語を元に，ブックマークされたページの内容に沿って「PHP」タグが付いたブックマークを分類する．これにより，このグループが持つ「PHP」

4.3　ブックマークされたページの内容に沿った分類

タグが付与されたブックマーク (A,B,C,D,...) があるとき，表 4.2 のように分類することが出来る．

表 4.1　「PHP」タグが付与されたブックマークの分類例

文書	ブックマーク
クラスタ 1	A G L O ...
クラスタ 2	D H J N ...
クラスタ 3	B E F M ...
クラスタ 4	C I K P ...

同様に,「歴史」「音楽」「食べ物」と，全てグループタグについて分類を行う．これらを全てのグループに行うことで，図 4.2 のように，全てのグループが持つブックマークを分類することが出来る．

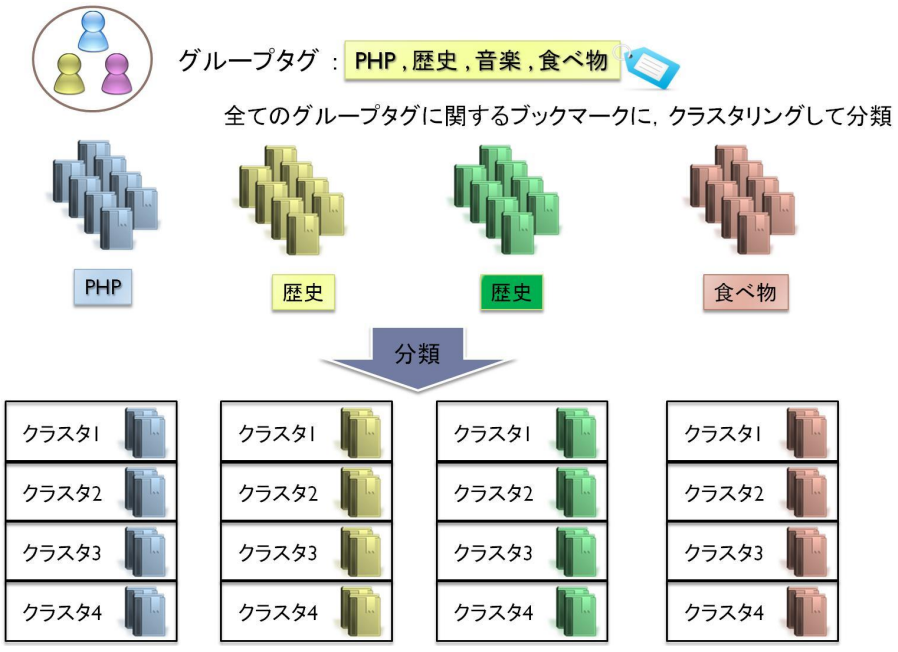

図 4.2　グループが持つブックマークの分類

4.4　ブックマークの推薦

クラスタリングによる分類により，利用者が属するグループが持つブックマークは，それぞれのグループタグ情報について，それぞれ 4 つに分類された．その中から利用者に最も合うものを選んでブックマークを推薦する．

利用者は提案システムを利用するとき，自身が属するグループに付与されたグループタグについて知りたい情報を選択する．利用者が属するグループに，グループタグ「PHP」「歴史」「音楽」「食べ物」が付与されているとする．利用者が「PHP」についての情報を得ようとした場合,「PHP」タグを選択する．このとき，前処理により利用者を含むそのグループが持つ「PHP」タグが付いたブックマークは，4 つのクラスタに分類されている．分類された 4 つのクラスタのブックマークを参照し，利用者が持つ「PHP」タグが付いたブックマークがどこに属するかを調べる．結果，利用者が持つ「PHP」タグが付いたブックマークが最も多く含まれるクラスタを，この利用者が属するクラスタとする．利用者が属するクラスタに分けられたブックマークの中で，この利用者以外の人がソーシャルブックマークに登録したブックマークをこの利用者に推薦することで，より利用者の嗜好に合わせた推薦をすることが出来る．その動作を図 4.3 に示す．

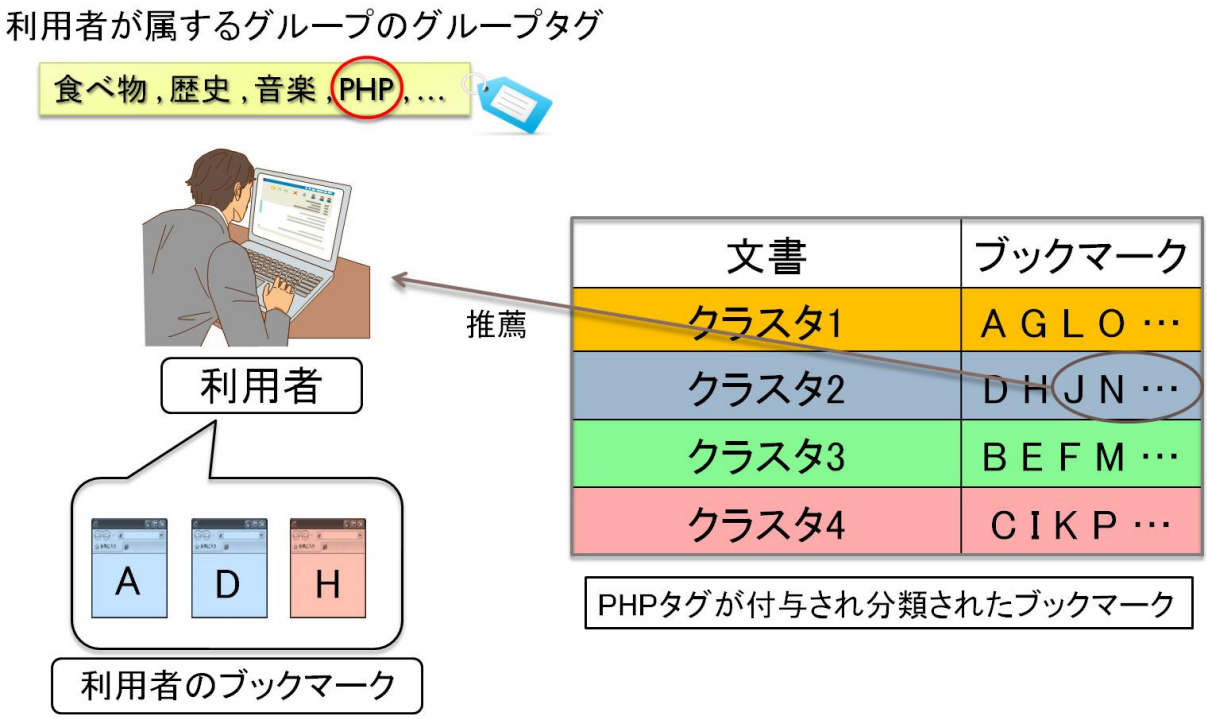

図 4.3　既存システムの実行画面

第 5 章

評価と考察

本章では，提案システムを評価し，その結果を考察する．ブックマークに対する分類が正しく行われているか，また 8 名の被験者が提案システムを活用した利用履歴から評価した．

5.1　分類の精度

あらかじめ 3 種類に分けられたページ，PHP の動作サンプル 112 ページ，PHP のプログラムが書かれたページ 97 件，PHP の概要・説明がなされたページ 91 件，計 300 件をクラスタリングを用いて分類を行った．クラスタリングによって分類されたページが本来分けられるべきクラスタに属する割合を適合率とし，評価式を式 (5.1) とした．

$$適合率 = \frac{自動分類により正しく分類されたページ数}{ある種類に本来分けられるべきページ数} * 100 \quad （\%） \tag{5.1}$$

とした．その結果，適合率は 93.3%となった．

5.2　利用履歴からの評価

8 名の被験者に提案システムを利用して貰った．それぞれの利用者が提案システムを利用した際に，推薦されるブックマークの上位 30 件を閲覧．有用だと判断したブックマークの数をカウントした．

そして，以下の評価式に代入して，有用であるブックマークの割合を算出した．評価式は式 (5.2) とした．

$$\text{有用率} = \frac{\text{有用だと判断したブックマークの数}}{30} * 100 \quad (\%) \tag{5.2}$$

である．8 名の有用率を平均した結果，81.6%という結果が得られた．

5.3 考察

提案システムにより，ブックマークされたページの内容に沿った分類が，適合率 93.3%となり，正しく行えていることが分かる．また，それによる利用者の嗜好に合わせたブックマークの推薦も，既存システムの 47.6%に比べて 81.6%と，高い結果を得ることができ，利用者に対して有用なページを推薦出来ていることが分かる．今後は，NMF を用いてクラスタリングをときの，クラスタの数の設定を変え，それぞれのグループタグに最も合う数を設定する手法を検討することで，さらに正確な分類を行うことができ，より利用者の嗜好に合ったブックマークが推薦できると考えられる．

第 6 章

おわりに

本論文では，組織で行われている情報共有の重要性について述べ，その中でもブックマークを用いた情報共有の利便性について述べた．さらに，ブックマークを用いた情報共有システムに対する検索支援システムとして既存方式に付いて述べた．既存方式では，タグによる人物の分類では，まだ推薦されるブックマークの有用率が低く，ブックマークされたページの内容に沿った分類が行われておらず，利用者の嗜好に合わせた推薦がされていないことが問題だと指摘した．その問題点に対して，単語抽出をした後，シソーラスを用いて表記揺れに対応し，tf-idf 法を用いて意味の無い単語の除外を行い，NMF と呼ばれるクラスタリング手法を用いて，ブックマークされたページの内容に沿って分類し，より利用者の嗜好に合わせたブックマークを推薦する手法を提案した．そして，提案方式の有用性を確かめるために評価実験を行った．その結果，シソーラス，tf-idf 法，NMF を用いた分類は正しく行うことが確認でき，さらに提案システムによるブックマークの推薦では，既存方式の有用性 47.6%を上回る，81.6%という結果が得られ，提案方式が有用であることが証明された．まとめとして，組織で使われているブックマークを用いた情報共有システムを用いるとき，より利用者に有用なブックマークの検索支援が行えるようになった．

今後，サービス化を行うときの問題点として，クラスタリングを行う頻度を検討する必要がある．ブックマークは日々増減していくものである．ブックマークが 1 件変化するたびにクラスタリングによる分類を行うと，組織内でソーシャルブックマークを利用している人数が多ければ多いほど，1 日に行うクラスタリングの処理回数が増え，サーバに負荷がかかりすぎてしまう．そこで，ブックマークが何件変化したときにクラスタリングを行う，または何日に 1 回クラスタリングを行うなど，クラスタリングを行う頻度を検討する必要がある．

謝辞

本研究の遂行と論文作成にあたって，言葉では言い表せないほどの御指導，御助言をいただきました高知工科大学フロンティア工学コース 清水明宏教授に心より感謝し厚く御礼申し上げます．本研究の副査を担当していただいた高知工科大学フロンティア工学コース 野中弘二教授，古沢 浩教授に深く御礼申し上げます．

また，提案システム実装，評価にご協力いただきました清水研究室の皆さまに心より感謝いたします．

最後に，有益な議論を交わしていただいた高知工科大学 清水研究室の関係者各位に深く感謝いたします．

参考文献

[1] ソーシャルブックマーク利用と導入のポイント - Enterprise 2.0 の現状 - ZDNet Japan, http://japan.zdnet.com/sp/feature/07sp0060/story/0,3800076669,20346851,00.htm, , 2007..

[2] The ISC Domain Survey — Internet Systems Consortium, http://www.isc.org/solutions/survey, , 2013

[3] Jesse Alpert and Nissan Hajaj. “ We knew the web was big... ”. Official Google Blog. 2008-07-25. http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html, 2007.

[4] Karger,D.R. and Quan,D., What would it mean to blog on the semantic web?, in Proc. Third International Semantic Web Conference, pp.214-228 ,2004.

[5] 毛受崇 吉川正俊 , ブックマークの時系列情報を利用したソーシャルブックマークにおける注目度予測, 電子情報通信学会 第 19 回データ工学ワークショップ, 2008

[6] 森 一聡, “ ブックマーク情報を用いた Web 検索支援システムの開発 ”, 高知工科大学フロンティア工学コース, 学士学位論文, 2010.

[7] Livedoor 研究用データセット 2008 年 12 月までのデータ, http://labs.edge.jp/datasets/, 2009.

[8] 形態素解析器「茶筌」 Ver.2.3.3, 奈良先端科学技術大学院大学 松本研究室 ,http://chasen.naist.jp/hiki/ChaSen/, 2003.

[9] 丸太要, 中村貞吾, ”文書類似度を考慮した NMF を用いた記事カテゴリ判定, ”九州工業大学情報工学部, 情報処理学会研究報告 2012.

[10] 大谷紀子, “ 情報検索におけるベクトル空間モデルの応用, ”武蔵野工業大学環境情報学部研究論文 pp.3-6, 2004.