

# 閲覧履歴に基づく情報検索の相互支援システムの開発

根本 潤† 遠山 元道†

†慶應義塾大学 理工学部 情報工学科

## 1 はじめに

近年、Web で発信される情報量は飛躍的に増加している。しかしながら、その情報量の多さが時に弊害となる場合がある。例えば、一人のユーザが閲覧可能な文書数には限りがあり、本来必要としている文書であっても発見することができなったり、見逃してしまったりということがしばしば起こりうる。

こうした状況を避けて必要な情報だけを効率的に得るため、本研究では小規模な組織やワークグループで利用する情報検索の相互支援システムを開発した。

## 2 情報検索の相互支援

本システムのアルゴリズムの概要について述べる。ここでは簡略化のために、推薦を行うユーザと推薦されるユーザの二人だけが検索を行う状況を想定する。なお、前者の立場のユーザをアドバイザ、後者の立場のユーザをレシーバとして説明する。

まず、システムではバックグラウンドの処理として次のことが随時行われている。

- ユーザが閲覧したページを履歴として保存する。
- TF・IDF 法により、各閲覧履歴の特徴キーワードを抽出し、特徴ベクトルを生成する。またそれにより、ユーザの興味ベクトルを生成する。

アルゴリズムの概要は以下の通りである。

1. 二人のユーザが自由に Web の閲覧を行う。
2. 閲覧中のページの特徴ベクトルを生成する。

Development of Mutual Support System for Information Retrieval based on Browsing History

NEMOTO Jun†, TOYAMA Motomichi†

†Department of Information and Computer Science, Faculty of Science and Technology, Keio University.

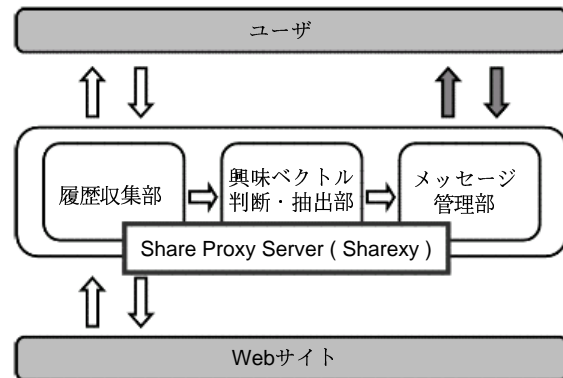


図 1: システム概要

3. アドバイザの閲覧中のページの特徴ベクトルとレシーバの興味ベクトルの類似度をコサイン相関値により求める。
4. 類似度が閾値を超えたら、システムはそのページが推薦候補であることをアドバイザに通知する。
5. アドバイザの主観的な判断により、レシーバにとって有益でありそうなら実際に推薦を行う。

## 3 システム概要

図 1 に、本システムの概要図を示す。システムは基本的に HTTP プロキシサーバとして実装されている。また、主として履歴を収集するモジュール、ユーザの興味・関心を抽出するモジュール、アドバイザやレシーバとのメッセージ管理を行うモジュールから構成される。

### 3.1 履歴収集部

履歴収集部ではプロキシサーバを通して閲覧されたページを各ユーザごとに保存していく。また、この履歴収集部でユーザが過去にどういったページを閲覧したかという一貫的な情報の管理も行う。

### 3.2 興味ベクトル判断・抽出部

興味ベクトル抽出部では履歴収集部で得られた各ユーザの閲覧履歴からユーザの興味ベクトルを生成する。これは、形態素解析システムの茶筌 [1] によって名詞に分類されたキーワードの出現頻度に基づいて生成する。

### 3.3 メッセージ管理部

メッセージ管理部にはアドバイザーやレシーバがシステム側と推薦や推薦候補の通知メッセージなどをやりとりするための機能が実装されている。メッセージのやりとりにはマイクロソフト社の MSN Messenger [2][3] を使用する。MSN Messenger は、インスタントメッセージのやりとりを行うソフトウェアで、オンライン中ならオンラインの相手に即座にメッセージを送ることができる。この機能を利用することにより、オンライン中のユーザに関しては即座に推薦を行うことが可能となり、検索の相互支援をより活発なものにする。なお、レシーバがオフラインの場合は同様の内容をメールで通知する仕様になっている。

### 3.4 利用例

本システムを利用するにあたっては、まず MSN Messenger の利用を開始しておく。続いてシステムにログインする。ログイン後は、通常通りのブラウジングを行う。すると、あるとき図 2 のようにシステムからメッセージを受け取る。これは閲覧中のページがレシーバへの推薦候補であることを示している。ここでシステム側に YES か NO で返事をする事で、実際に推薦を行うことができる。

## 4 考察

本システムを利用することにより、単独で Web の閲覧・検索を行うときよりも効率的に情報収集可能になることを図 3 に示す。

注目すべきは、濃く塗りつぶされた部分である。この部分はレシーバの興味・関心はあるが、まだ閲覧していない文書で、かつアドバイザーの閲覧した文書である。つまり、この部分が本システムによりレシーバが追加的に得られる文書集合である。この概念は複数のユーザについても適用することができるため、本システムにより小規模な組織内で相互補完的かつ効率的な情報検索が可能になると考えられる。



図 2: 推薦候補を通知された状態

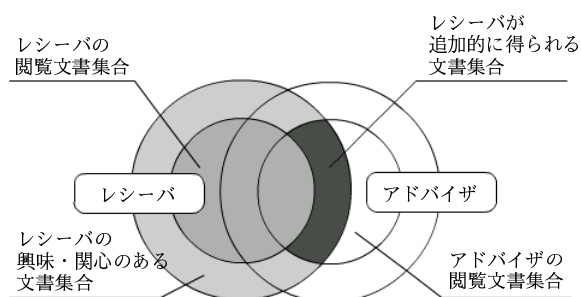


図 3: 本システムで追加的に得られる文書

## 5 まとめ

本研究では、閲覧履歴に基づく情報検索の相互支援システムを開発した。これにより、他のユーザとの連携の中で、これまで大量の情報の中に埋もれていた未発見の有益な情報を効率的に獲得可能となった。今後は、推薦を受けたときのレシーバの評価を用いて、興味ベクトルを修正していく機能などを追加し、推薦の精度を向上させることを検討している。

## 参考文献

- [1] 松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 高岡 一馬, 浅原 正幸 “形態素解析システム『茶筌』 version 2.3.3 使用説明書” 奈良先端科学技術大学 情報科学研究科 自然言語処理講座, 2003
- [2] MSN Messenger: <http://messenger.msn.co.jp/>
- [3] MSN Messenger Protocol: <http://www.hypothetic.org/docs/msn/index.php>