

Micro media を用いたパーソナライズ情報収集支援システムの提案

澤田 瑞穂^{*1} 戸田 健^{*1} 尾崎 信耶^{*2}

^{*1} 日本大学理工学部電気工学科

^{*2} 日本大学大学院理工学研究科電気工学専攻

csmi08805@g.nihon-u.ac.jp

概要

WEB 上の情報は拡大する一方であり、従来の検索方法では莫大な情報を精査するのに時間と労力を要する。本研究では、ユーザが発信する Micro media 情報を無意識に利用することで、短時間に所望の情報を得ることができる検索システムを検討する。提案システムでは、ユーザによって Twitter、Blog や SNS をはじめとする Micro media に投稿されたユーザに関する情報から、名詞だけでなく動詞に注目したテキストマイニングを行うことでユーザの状況を把握する事により検索語を抽出し、検索及び検索結果のパーソナライズを実現する。

1. はじめに

WEB が普及した今日では、ソーシャルネットワークサービス(Social Network Service : SNS)や Blog、Twitter 等のマイクロメディアを用いた一般ユーザによる情報発信が容易なものとなり、今まではあまり重要視されていなかった一般ユーザの発信する情報が重要な情報源として利用されるようになってきている。例として、中東・北アフリカをはじめとする大規模な反政府デモを支えたのは、WEB を通じて発信された市民からの情報である。また、日本においても東北・関東大震災発生直後では、情報が錯綜する中でマイクロメディアが情報収集の手段としての重要な役割を担っていた。

こうした一般ユーザによる情報も重要視されてきた状況で、情報取得の面から WEB を見ると、膨大な数の一般ユーザによる情報も有用になっており、WEB 検索エンジンを利用したとき一度に取得される情報が膨大なものになっている。また、従来の WEB 検索エンジンでは検索結果の表示に関して、単語と WEB ページ内容の関連性や WEB ページの被リンク数による WEB ページの重要度を一般化し、重要度のランキングに則して表示されている。マイクロソフト株式会社の調査^[1]では、30代~40代のユーザのうち約 85%が「検索結果から適切だと思われる結果を探し、検索結果サイトを順にクリックして探すことに時間をかけている」と答えている。本論文ではこのような時間をかけて情報を精査する従来の検索方法を能動的な検索方法と呼ぶことにする。また、同調査結果から、30代~40代のユーザのうち約 60%が「検索したキーワードで欲しい情報が見つからない」と答えている。これは明確なキーワードが見つからずに所望の情報が得

られないため、曖昧な所望の情報を検索語に反映させることは一般ユーザには難しい事だと考えられる。この能動的な検索方法で得られる膨大な情報や、一般化された検索結果の配列は、ある程度情報を精査する時間があり検索語が明確な場合には有用であるが、即座に情報が欲しい場合や曖昧な検索語には不向きである事がわかり、曖昧な目的の検索にかかる時間と労力はユーザが「情報を得たい」と考えた段階での「情報に対する行動意欲」を削ぐと考える。

そこで、本提案システムでは以上の観点を踏まえ、日常生活動作の一部となっている WEB 検索の時間をかけないで即座に有用な情報の取得を実現することでユーザの行動意欲の明確化を実現することを目的とする。その上で、能動的な検索方法に対し、受動的な検索方法とそれを実現するシステムを提案する。受動的な検索方法とは、ユーザが能動的な検索をすることなく、自分の発言から所望の情報をコンピュータから提供してもらった検索方法である。例えば、通勤途中にふと、「(明確な目的は無いものの)旅行に行きたい。」と思いソーシャルメディア上で「旅行に行きたい。」と発言したとする。今までは膨大な情報の中から自分のイメージした旅行先を絞り込んでいかなければ、明確な意欲に繋がらない場合や、旅行に行く事をあきらめたかもしれないが、本システムによって無意識に思っていた事と一致する情報が手に入れば、無意識のうちの曖昧な意欲をより明確化でき、ユーザの行動の機会を増やすことにつながる。また、既存のレコメンドサービス等で行う所望の情報元の登録作業を排除することで、パーソナル情報の推移にも十分対応できると考えられる。

提案システムでは、ユーザによってマイクロメディアに投稿された情報から、名詞だけでなく動詞に注目したテキストマイニングを行うことでユーザの状況を把握する事により検索語を抽出し、同時に検索結果の推測を行い、検索及び検索結果のパーソナライズを実現する。

また、インタフェースについては、ユーザは一切検索には関与せず、自分の発信情報を軸に情報が集まってくる、という形をとった。

2. 関連研究

2.1 WEB 検索の現状について

WEB 検索方法として、「検索したい情報の名前(単語)を考える→単語での検索をする→膨大な情報からページを選択し、所望の情報を得る」という方法が挙げられる。この方法は一般化しており、また検索される側であるWEB ページ作成者もそのような検索に対して単語での検索に対応したページ作りをしている。しかし、単語の意味は、用いられる状況に応じて変化するものである。また、ユーザの知識が乏しい場合もあり、所望の情報を単語で表現できない場合や、第一章で示したように、現在のWEB の状況はリアルタイムで情報が生まれている状況であるため、まだ名前の定まらない情報もある。

こうした状況に対し、「セマンティック WEB」^[2]の導入が検討されている。セマンティック WEB とは、コンピュータ側で情報の意味(メタデータ)を理解し、その意味から多角的に情報を取り出す事ができる、いわば逆引き辞書のような考え方である。例えば、検索エンジンに「サイパンに行きたい。」と入力すれば、従来方法の単語検索では「サイパンへ行きたい」または「サイパン+行く」という語を含む情報が結果として得られていたが、セマンティック WEB では、サイパンの情報やサイパンへの航空機の搭乗券の情報、サイパンの天気などといった、「サイパンへ行きたい」という語の意味を理解して、そこから抽出されるサイパンへ行くという事の意味から推測された情報を得る事ができる。

2.2 情報レコメンドサービスについて

現在、WEB 上において様々なレコメンドサービスの開発が進んでいる。各種商品購買サイトにおける商品の表示履歴や購入履歴から同様な購入履歴を持つユーザの履歴に残った商品を薦める、「商品推奨サービス」^[3]をはじめ、Blog 内広告の内容を Blog の内容から自動的に選択する Yahoo Japan の「インタレストマッチ」^[4]が上げられる。また携帯端末上での情報レコメンドサービスとして、NTT DOCOMO による「i コンシェル」^[5]がある。しかし、これらのサービスでは、過去のユーザの所望の情報から提供情報を選択し提供するために、情報の新規性が失われている場合が考えられる。また、ユーザによる情報提供サイトの登録によって新規情報を得る事ができるが、ユーザ自身の所望の情報は常に変化しているものと考え

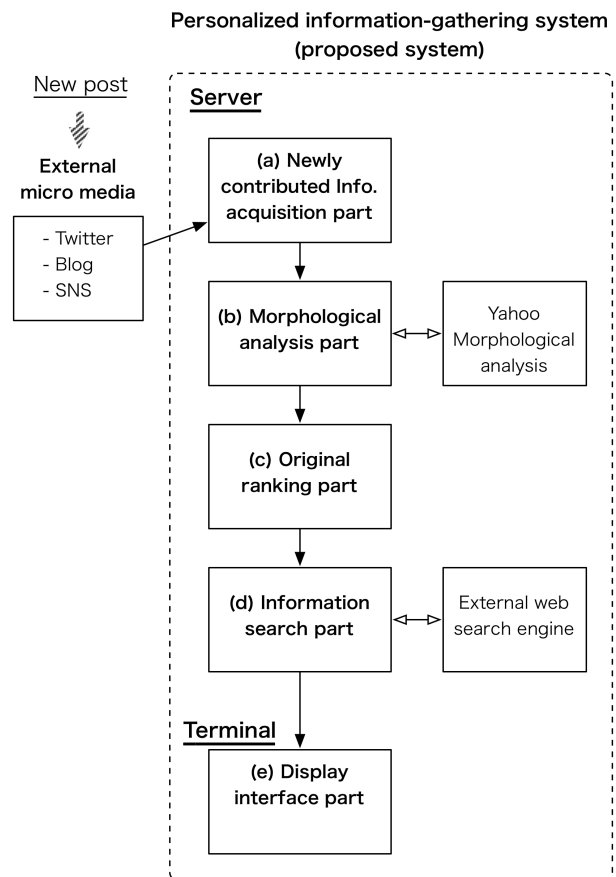


図1 提案システムの機能構成図

られ、登録した情報配信が常に所望のものであるという事はない。

2.3 インタフェースについて

インタフェースについて、ユーザの活動の合間時間を利用して情報を獲得する「眺めるインタフェース」^[6]を渡邊らは提案している。また、システムが自動的に情報を次々と掲示する「掲示型インタフェース」^[7]を水口らは提案している。

3. 提案システム

3.1 機能構成

提案システムの機能構成を図1に示す。なお、提案システムでは、ユーザの日常動作を重要視するので、普段から使用しているソーシャルメディアを従来どおり使用する。提案システムの機能構成は、図1よりA. 外部ソーシャルメディアシステム、B. 情報集支援システムに大別される。ユーザは、あらかじめ端末上から自身が使用しているソーシャルメディアのURLをB. 情報収集支援システムに登録しておく。次に、システム構成要素の概要を説明する。

3.1.1 外部ソーシャルメディア

Twitter や Blog, SNS 等従来ユーザが使用していたものを従来通り用いる。

3.1.2 パーソナライズ情報収集支援システム

サーバ上のパーソナライズ情報収集支援システムは以

下(a)~(d)の4つの部分から成る。

(a) 新着投稿情報取得部

登録されている外部ソーシャルメディアにおいて、ユーザが新規投稿をすると、新着投稿情報取得部が新着投稿の内容の文字情報を取得する。

(b) 形態素解析部

(a)で取得した文字情報を形態素解析し、名詞部と動詞部のみ抽出する。Yahoo 形態素解析^[8]を利用した。

(c) パーソナライズランキング部

名詞部と動詞部からシステム独自の重み付けを行い、ランキング化する。このランキングによって検索語の生成と検索結果のパーソナライズを行う。本機能は、本システムの重要部であるので、次節で詳細を述べる。

(d) 情報検索部

Google、Yahoo などの外部検索サイトを用いる。

(e) 表示インタフェース部

表示部は端末上の実装し、サーバから送られてきた検索結果を表示する。この結果表示方法に関しては、本提案システムの重要な点であるため 3.4 節にて詳細を述べる。

3.2 動作の流れ

前節で示したシステムの実施例を述べる。ユーザがあらかじめシステムに登録した外部サイトに記事を投稿するとサーバ上のが情報の取得を判断し、外部ソーシャルメディアから新着記事の文字情報を取得、形態素解析処理をする。処理結果より得られた名詞+動詞に着目し、重み付けを行い、検索語の抽出を行う。検索語と過去の関連語を元に外部 WEB 検索エンジンにて情報検索をし、得られた結果を再びパーソナライズ情報収集支援システムで処理し、上位数件を端末に送り端末上に表示する。

3.3 パーソナライズランキング部

本研究では、ユーザが受動的な検索をすることが課題である。この受動的な検索の実現においてパーソナライズ部は最も重要である。なお、この点についてユーザの関与はほとんど無く、すべてコンピュータが処理する。これは、ユーザに検索するという動作を意識させないためである。パーソナライズランキングは、パーソナライズ機能としての働きをする。このランキングでは独自にいくつかの動詞をあらかじめ選出しておき、重みを与えておく。ユーザのソーシャルメディア上での発言から抽出された動詞についてあらかじめ選出されている動詞とマッチしたときに、マッチした抽出された動詞に伴って抽出されている名詞に選出されている動詞に与えられている重みを与えることで、その重みに応じて名詞のランキングを作るというものである。例を図2に示す。ユーザが〇〇コーヒーが好きだ。と発言すると、システムが

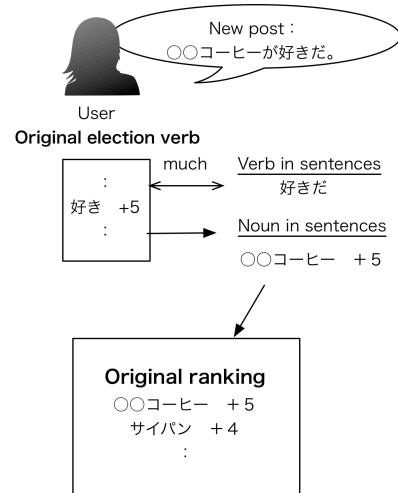


図2 独自ランキング

発言から動詞+名詞を抽出する。あらかじめ独自に選出してある5の重みを持った「好き」という動詞と、発言の「好きだ」という動詞がマッチしたことで、抽出された名詞「〇〇コーヒー」は5の重みを与えられ、名詞情報がランキングとして残る。このランキングを元に検索語の決定や検索結果のフィルタリングを行う。また、このシステムはユーザのリアルタイムの発言で変化するランキングである。このため、常に変化するパーソナル情報に対応することができると考えられる。

3.4 表示インタフェース部

本研究のインタフェースについては、受動的な検索の実現のためにユーザがふとした瞬間に、端末上に表示された偶然に興味と一致する情報を見ることで情報選択への意欲を高めるという狙いがある。通勤・通学や帰宅途中での移動中の利用や待ち時間での利用を想定するので、動作時間が不定である。このため、常に端末画面上で情報が更新されており、システムの起動や終了に関してはユーザの関与が無い様にする。また、情報の表示方法については動きを持たせる事でユーザへある程度の注意を引くようにする。

4. 今後の評価計画

提案システムを実際に試作し、実環境における評価を通して本用途における提案システムの有効性について検証を行う。

4.1 評価方法

実験システムの構成を図3に示す。構成は、外部マイクロメディアサイトとパーソナライズ情報収集支援システム(提案システム)からなる。パーソナライズ情報収集支援システムはサーバ上と各端末上にあり、サーバ上でパーソナライズ処理をして、端末上には検索結果の表示部のみを搭載する。これは、端末上での各機能の動作は端末に大きな負担かけると考えられるためである。以下に、検証をしていく上での評価の必要な点について述

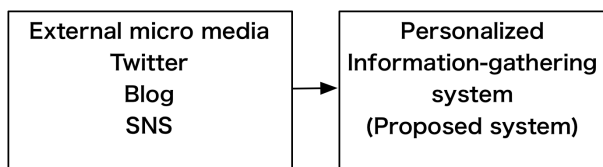


図3 システム構成図

べる。

4.2 パーソナライズランキング部の評価

パーソナライズランキングの目的は、ユーザのパーソナル情報を取得することであるため、ユーザの発言情報とマッチさせる独自の選出単語について検討が必要である。本システムを用いての精度の高いパーソナライズの実現には、ソーシャルメディアを用いてリアルタイムに取得した情報から、従来のWEB検索では知る事のできなかったユーザの状況をより正確に把握する事が求められる。選出単語から、何程の状況判断ができるのか、ランキングに基づく検索結果と実際にユーザが所望した情報との検索結果の的中率について評価の必要がある。

4.3 表示インタフェース部の評価

検索結果の表示については最終的にユーザが検索結果の情報を見て行動意欲を持つ事が求められ、この点はシステムの有効性について最も重要なものである。このため、システムを用いたユーザが結果の表示方法による行動意欲への影響力を調べる必要がある。

表示方法については、

- ・ 検索結果の概要を表示する方法
- ・ 検索結果のURLを表示する方法
- ・ 検索結果サイトの画像を表示する方法

等、いくつか上げられる。

5. おわりに

本稿では、マイクロメディアを用いたユーザのリアルタイムな投稿から独自に検索語のパーソナライズを行いユーザに意識させない形でWEB検索をすることで、時間と労力がかかるようになってきた従来のWEB検索の方法を解決する方法を提案した。今後は、実際にシステムを試作し、実験を通してシステムの最適化を図る。また、今後はシステムを発展させて発言情報に付与されるGPS位置情報からユーザの行動を分析することや、発言時間に合わせた情報を提供することも検討し、より高度なパーソナライズ情報収集支援システムを提案する予定である。

参考文献

[1] マイクロソフト株式会社:検索エンジン利用に関するアンケート;
<http://beta.keyword.jp.msn.com/bing/summary.htm>
 (2010).

[2] 林:セマンティック Webと言語資源・言語技術; IPSJ Magazine Vol.48 No.8 Aug. 2007,pp.857-863 (2007).
 [3] Amazon.com:おすすめ商品;
<http://www.amazon.co.jp/gp/help/customer/display.html?nodeId=779360> (2011)
 [4] 株式会社エヌ・ティ・ティ・ドコモ:i コンシェル;
<http://www.nttdocomo.co.jp/service/customize/iconcier/>
 (2011)
 [5] ヤフー株式会社:インタレストマッチ;
<http://listing.yahoo.co.jp/service/int/index.html?o=IM0024> (2011)
 [6] 渡邊,安村:Memorium:眺めるインタフェースの提案とその試作;インタラクティブシステムとソフトウェアに関するワークショップ(WISS2002)論文集,pp.99-104 (2002).
 [7] 水口,浦野:揭示型インタフェースによる情報検索システム;インタラクティブシステムとソフトウェア VI WISS'98,pp.159-164,近代科学社(1998).
 [8] ヤフー株式会社:テキスト解析 Web API 日本語形態素解析; <http://developer.yahoo.co.jp/webapi/jlp/> (2011)