

ユーザの嗜好に基づいた RSS ニュースリーダーに関する基礎研究

吉田博哉[†] 田中成典[‡] 杉之原亮[†]

関西大学大学院[†] 関西大学総合情報学部[‡]

1. はじめに

近年，インターネットの普及に伴い，誰もが Web から様々な情報を取得できるようになった．特に，ニュースサイトなどの更新頻度の高い情報を提供する Web サイトが増加[1]し，Web からの情報収集はさらに活発になっている．しかし，Web の膨大な情報の中から目的の情報を取得するためには，多大な時間的コストがかかるという問題[2]がある．そこで，Web サイトの更新情報を配信，取得できる RSS (RDF Site Summary) が注目されている．RSS とは，更新情報のタイトル，要約や更新時間などを構造化して記述する XML ベースのフォーマットであり，Web で公開されている RSS の情報は，RSS リーダにより容易に取得できる．これにより，Web サイトの更新情報を効率的に取得できるようになった．しかし，ニュースサイトなどの RSS による情報配信を行っている Web サイトの急増により，RSS の情報は氾濫し，ユーザの情報検索効率が著しく低下している．既存システムでは，RSS の記事からユーザの興味を学習し，興味のある記事を優先的に提供する RSS リーダが開発されている．しかし，既存システムでは，閲覧記事から抽出した特徴語の出現回数のみで興味を推定しているため，興味の学習効率が悪いという問題がある．そこで，本研究では，記事閲覧の時間からユーザにとって興味のある特徴語の抽出[3][4]や，記事閲覧の順番[5]から特徴語の興味度の更新を行い，ユーザの興味を迅速に学習することで，取得した RSS の記事を興味度の高い順番に並べ替えてユーザに提供するシステムを考案する．

2. システムの概要

本システムでは，図 1 に示すように，1) 興味記事解析機能，2) 興味度付加機能，3) 興味学習機能，4) 記事の興味度算出機能の 4 つ

の機能により構成される．

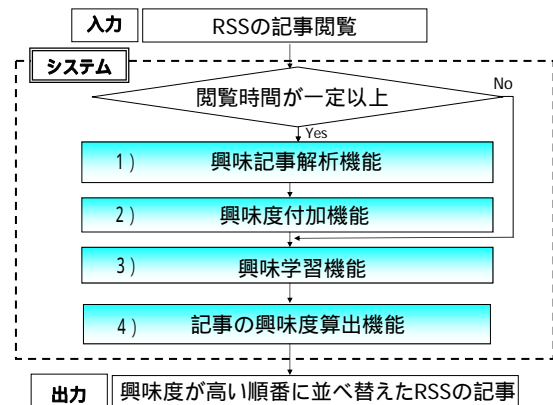


図1 システムの流れ

2.1 興味記事解析機能

興味記事解析機能では，ユーザの興味のある記事を取得し，記事の特徴語を抽出[3]する．まず，興味のある記事を，記事の閲覧時間の長さから興味の有無を判定し，取得する．次に，取得した RSS の記事から，記事のタイトルと要約そして閲覧した記事の本文を抽出する．最後に，形態素解析を用いて，記事のタイトル，要約と閲覧した記事の本文から，名詞と未知語を特徴語として抽出[4]する．

2.2 興味度付加機能

興味度付加機能では，興味記事解析機能で抽出した特徴語と特徴語同士の共起に対して，ユーザの興味度として重みを付加する．特徴語への重みは，「記事の本文」，「RSS の記事の要約」，「RSS の記事のタイトル」から抽出した特徴語の順で大きくする．

2.3 興味学習機能

興味学習機能では，ユーザが閲覧した RSS の記事の順番から特徴語の興味度を更新する．ユーザが閲覧した記事の順番は，ユーザの興味に大きく影響する．そのため，本システムでは，記事の閲覧順位に影響する要因を推定し，ユーザの興味をマイニングする「順位付け文書からの影響因子マイニング[5]」を使用することにより，記事の閲覧順位の情報からユーザの興味を抽出する．

Fundamental Research on RSS News Reader Based on User's Preference

[†]Hiroya Yoshida

Graduate School of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka, 569-1095, Japan

[‡]Shigenori Tanaka, Ryo Suginoara

Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka, 569-1095, Japan

2.4 記事の興味度算出機能

記事の興味度算出機能では、興味度付加機能と興味学習機能により付加した興味度から、取得した RSS の記事に対してユーザの興味度を算出する。記事への興味度は、RSS の記事の特徴語と特徴語同士の共起に付加されている重みの合計によって算出する。

3. システムの実証実験と考察

本システムの実行結果を図 2 に示す。実証実験では、本研究で開発したシステムの有効性を検証するために、既存の興味学習機能が備わった RSS リーダを使用した場合との更新情報の検索効率における比較実験を行った。

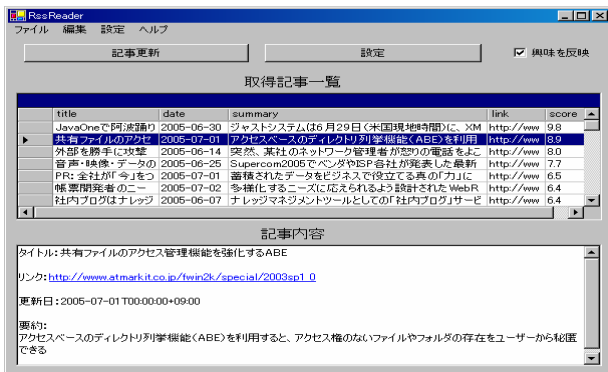


図 2 実行結果

3.1 実証実験

実証実験として、日常的に RSS リーダを使用している被験者 10 名を対象に同一のニュースサイトから RSS の記事を取得する作業をそれぞれ 20 回行った。次に、本システムの評価として、5 段階のアンケートを実施した。被験者には、RSS の記事を取得して閲覧するまでの手順の単純さを評価する「操作性」、ユーザの興味が記事に反映された期間を評価する「即時性」、ユーザの興味のある記事の検索の容易さを評価する「利便性」、ユーザの興味のある記事の取得におけるシステムの有用性を評価する「有用性」の 4 つのアンケート項目を回答してもらうことにより本システムを評価した。

3.2 結果と考察

アンケート結果を基に算出した各評価項目の平均値と今回のアンケート結果に有意性があるかを調べるために行った U 検定の結果を表 1 に示す。U 検定とは、正規分布が仮定できない 2 標本の数量の平均値の差を検出する検定である。U 検定の結果、5%の有意水準で「即時性」、「利便性」、「有用性」、の 3 項目で本システムの有意性を確認できたが、「操作性」では、有意性を確認できなかった。その原因としては、

本システムは、興味のある記事を効率的に提供することが主な目的であり、操作性に関しては従来のシステムと差がなかったためであると考えられる。

表 1 実験結果

評価項目	本システム (平均値)	既存の RSS リーダ (平均値)	U 検定 (有意確率)
操作性	2.90	3.00	0.720
即時性	4.10	2.90	0.004
利便性	3.60	2.60	0.021
有用性	3.70	2.80	0.040

4. おわりに

本研究では、RSS で取得した記事をユーザの興味の高い順番に並べ替えて提供する手法を考案した。実証実験では、本システムを利用することで、既存の RSS リーダよりも、被験者の興味に合った記事を迅速かつ優先的に提供することができた。この実験結果から、本研究で提案する手法の有効性を実証した。しかし、本システムでは、記事から抽出した特徴語の類義語を考慮していなかったため、同じ意味の語でも別の語として判別されてしまい、ユーザの興味が記事に反映できないという現象が生じた。今後の課題は、特徴語の類義語も考慮することで、システムの精度向上を図る。

参考文献

- [1]Bing Tan, Schubert Foo, Siu Cheung Hui: Web information monitoring: an analysis of Web page updates, Online Information Review, Vol.25, No.1, pp.6-19, 2001.2.
- [2]土方嘉徳: 情報推薦・情報フィルタリングのためのユーザプロファイリング技術, 人工知能学会誌, 人工知能学会, Vol.19, No.3, pp.365-372, 2004.5.
- [3]中島伸介, 黒田慎介, 田中克己: 閲覧履歴を反映したコンテキスト依存型 Web ブックマーク, 情報処理学会論文誌, 情報処理学会, Vol.43, No.14, pp.23-36, 2002.6.
- [4]Yutaka Matsuo, Hayato Fukuta, Mitsuru Ishizuka: Browsing Support by Highlighting Keywords based on a User's Browsing History, Proceedings of the IEEE International Conference Systems Man and Cybernetics, IEEE, Vol.2002, No.6, pp.174-179, 2002.10.
- [5]沢井康孝, 峠泰成, 山本和英: 順位付け文書からの影響因子マイニング, 情報処理学会研究報告, 情報処理学会, Vol.2004, No.93, pp.169-176, 2004.9.