ブロガーの熟知度に基づいたブログランキング方式の提案

中島　伸介† 　稲垣　陽一†† 　草野　奉章††

† 奈良先端科学技術大学院大学 情報科学研究科　〒 630–0101 奈良県生駒市高山町 8916-5
†† 株式会社きざしカンパニー　〒 103–0015 東京都中央区日本橋箱崎町 24-1 日本橋箱崎ビル 2F
E-mail: †shin@is.naist.jp, ††{inagaki,kusano}@kizasi.jp

あらまし　近年，ブログや SNS 等の CGM と呼ばれるコンテンツが数多く配信されるようになり，これらに対する検索要求も高まっている．本研究では，ブロガーが過去に投稿したエントリに含まれるキーワードの頻度から，そのキーワードが表す話題に対するブロガーのマニア度を算出し，これに基づいてブログエントリのランキングを算出しようとするマニア指向ブログランキング方式を提案する．また，実装したプロトタイプシステムに基づいた評価実験を行ったので報告する．本手法は単にランキングの尺度を与えるのみならず，ある話題に対して視点が異なるエントリの呈示が可能になると考えている．なお，今後の研究の方向性についても述べる．

キーワード　ブログ検索，ブログランキング，Ｗｅｂマイニング

Introduction of Blog Ranking Method Based on Bloggers' Knowledge Level

Shinsuke NAKAJIMA†, Yoichi INAGAKI††, and Tomoaki KUSANO††

† Graduate School of Information Science, Nara Institute of Science and Technology
†† kizasi Company, Inc.
E-mail: †shin@is.naist.jp, ††{inagaki,kusano}@kizasi.jp

Abstract　Recently, search requests for CGM contents such as blog and SNS are increasing. Thus, we propose an enthusiast-oriented ranking method for blog search engine by means of the enthusiast score of each blogger calculated based on term frequency in his/her past blog entries. We evaluate our proposed method using the prototype system as well. Our method achieves not only providing a blog ranking but also offering a choice between different ranking lists with an each viewpoint. Moreover, we describe the future work of our research.

Key words　blog search, blog ranking, Web mining

1. はじめに

近年，ブログや SNS 等の CGM と呼ばれるコンテンツが数多く配信されるようになり，ブログ検索やそのランキングに対する要求が高まっている．ブログコンテンツの魅力の一つは，その即時性である．したがって，ブログランキングでは，エントリ投稿直後の短時間においてランキングを算出する必要がある．Web 検索エンジンとして最も大きな成功を収めた Google が採用している PageRank [1] は，リンク構造に基づいてランキングを計算するものであるが，投稿直後のブログエントリには第三者からのリンクは存在しないため，このようなリンク構造解析をブログランキングに適用することは適切とはいえない．

最近では，ブログに特化した検索エンジンも数多く運用されている [2] [3]．これらの検索エンジンが採用しているランキング方式には，ブログエントリに対するランキングと，ブログサイトに対するランキングが存在する．

ブログエントリに対する従来のランキング手法としては，投稿時刻の新着順や，キーワードとの関連性に基づくランキングが採用されている．こちらは，Google ブログ検索 [4] 等が該当する．これらのランキングでは，コンテンツの価値を評価することができないことが問題である．

ブログサイトに対する従来のランキング手法としては，サイト全体に対するリンク数，アクセス数，投票等によるランキングが採用されている．Technorati [5] 等が該当する．これらのランキングでは，価値の高いエントリを投稿しているサイトを発見することができても，そのサイトに最新エントリが存在するとは限らない．したがって，最新エントリを取得できる保証がないことが問題である．

このような状況を受けて，ブログの特長である即時性を失うことのない，効果的なランキング方式が望まれている．

ブログの特徴として,“ あるブログサイトの内の全てのエントリは，原則として１人のブロガー（ブログ著者）によって書かれたものである ”というものがある．したがって，過去のエントリの履歴を見れば，そのブロガーの特性を解析して把握することも可能である．逆にそのブロガーの特性を把握することができれば，そのブロガーが新たに投稿したエントリの価値を推定することも可能である．例えば,“ Java ”に詳しいブロガーが書いた“ Java ”に関するエントリと，大学で初めて“ Java ”の講義を受けた学生が書いた“ Java ”に関するエントリとでは，前者の方がその専門度が高いと容易に推測できる．

したがって，ブロガーが過去に投稿したエントリに含まれる“ ある話題を表すキーワード ”の頻度から，そのキーワードが表す話題に対するブロガーの熟知度（マニア度）を算出し，これに基づいてブログエントリのランキングを算出しようとするブログランキング方式を検討する．

なお，本手法はランキングそのものの価値に加えて，複数の尺度に基づいたランキングをユーザに提示することが可能である．例えば,“ Java ”に関するエントリを検索した際に,“ Java ”そのものに対する熟知度に基づいたランキングだけでなく,“ プログラミング ”,“ 教育方法 ”,“ 試験対策 ”等に関する熟知度に基づいた複数のランキングを提示できる可能性がある．すなわち，単にシステムが与える一意のランキングからユーザが選択するのではなく，検索対象に対する複数の視点の中から，ユーザ自身が目的のものを選択することができるのである．

熟知度という概念に関しては竹原らの研究 [6] においても紹介されているが，熟知度判定の対象となる話題を表すキーワードの取得方法や，検索対象に対する複数の視点をユーザが選択可能にするような手法等，に関して考慮されておらず実用レベルに達しているものではない．

本稿では，ユーザに対して複数の視点に基づくランキングの提示が可能な，ブロガーの熟知度に基づいたブログランキング方式，すなわちマニア指向ブログランキング方式を提案すると共に，実装したプロトタイプシステムに基づいた評価実験を行ったので報告する．

以下，本論文の構成を示す．２節では，マニア指向ブログランキング方式について述べる．３節では，システムの実装およびについて述べる．４節では，今後の課題について述べる．５節では，まとめと今後の方向性について述べる．

2. マニア指向ブログランキング方式

本節では，ブロガーの熟知度に基づいたブログランキング，すなわちマニア指向ブログランキングの処理手順について説明する．

2.1 マニア候補辞書の作成

はじめに，ブロガーが記述するキーワードが，ある話題に関するマニア名として適切かどうかを判別する必要がある．例えば,“ ラーメン ”,“ タイガース ”,“ 鉄道 ”等は，その話題に関して熟知しているブロガーが存在すると予想でき，マニア名として適切である．一方,“ 最近 ”,“ 単純 ”,“ 効果 ”等の一般的な語句は，いわゆるマニアが存在するような話題としては適切とはいえない．そこで本研究では，事前にマニア候補辞書を作成することにした．以下に，マニア候補辞書の作成手順を示す．

（1） 「マニア」というキーワードで Web 検索を行い，検索結果のテキストの中から「マニア」の直前の名詞句をピックアップして，その頻度を計算する．(例えば,“ ラーメンマニア ”の場合は，語句“ ラーメン ”をピックアップする．)

（2） この頻度順に整列した名詞句のうち，頻度が高いものをリストアップする．

（3） 「ファン」「フリーク」に対しても，同様の作業を行う．

（4） 「マニア」「ファン」「フリーク」にて，リストアップした語のうち，重複を除去すると共に，人間の判断で不適切と思われる語句を除去した上で 辞書に登録する．

リストアップする際の頻度の閾値に関しては，低くしすぎると不適切な語句を数多く含めてしまう恐れがある．反対に高すぎると，十分な数のマニア名をリストアップすることができなくなる．もちろん，閾値の最適化を行ったとしても完全自動化は難しく，実用段階においては，他の方法を複合的に採用することや，人手による追加・削除の作業が必要である．しかしながら，本節で述べる手法を適用することにより，人的コストを大きく削減することが期待できる．

2.2 マニアブロガーの認定

本節では，各ブロガーがどのような話題のマニアであるのかを判定する方法について説明する．基本的には,“「マニア名を表すキーワード」を含むエントリを，長期間においてコンスタントに投稿しているブロガーをそのキーワードが表すマニアブロガーと認定する ”という方針に従う．ここで，マニアブロガーの認定基準として，以下の９個の条件を挙げた．

条件１：１件以上／週のエントリ投稿を３ヶ月間継続.
条件２：５件以上／月のエントリ投稿を３ヶ月間継続.
条件３：20 件以上／３ヶ月間のエントリ投稿
条件４：１件以上／ 10 日間のエントリ投稿を２ヶ月間継続.
条件５：３件以上／２週間のエントリ投稿を２ヶ月間継続.
条件６：15 件以上／２ヶ月間のエントリ投稿
条件７：１件以上／１週間のエントリ投稿を１ヶ月間継続.
条件８：３件以上／２週間のエントリ投稿を１ヶ月間継続.
条件９：８件以上／１ヶ月間のエントリ投稿

ただし，最終的に上記全てを採用しようとするのではなく，各条件によって認定されたマニアブロガーを検証し，条件の修正や追加を含め，最終的に採用する条件を決定する．また，この認定基準は対象となる話題（マニア名）によっても調整する必要があると考えられる．話題毎の基準の調整に関しては，今後の検討課題とする．

2.3 各ブロガーのマニアスコアの算出

マニアブロガーとして認定されたブロガーの熟知度を表すマニアスコアの算出方法を説明する．基本的な考え方としては，対象マニア名に関連する話題を含んだエントリの投稿数に基づいて算出する．

投稿したエントリと対象マニア名の関連度を計算するために，

マニア候補辞書に載っている全ての語句に対して，その関連語をデータベースに格納する．この関連語は，対象マニア名との共起度が高いもののみを格納した．今回は暫定的に，共起度上位 100 語とした．

ここで，対象マニア名 M に対する，あるブログエントリの関連度スコアを $escore_M$ とすると，

$$escore_M = \sum_{j=1}^{n} (w_j \cdot C_j \cdot E_j) \quad (1)$$

と表すことができる．ただし，w_j は共起度順位 j に基づく重み，C_j は共起度順位 j の関連語の共起度である．E_j は共起度順位 j の関連語が当該エントリ内に存在するかどうかを表現する変数であり，存在する場合 1，存在しない場合 0 の値をとる．また，n は関連度算出において考慮する関連語の数であり，今回は $n = 100$ である．

次に，対象マニア名 M に対するブロガー B のマニアスコアを $mscore_M(B)$ とすると，

$$mscore_M(B) = \frac{log(m)}{m} \cdot \sum_{i=1}^{m} escore_M(b_i) \quad (2)$$

と表すことができる．ただし，b_i はブロガー B が投稿した対象マニア名に関して書かれたエントリであり，$escore_M(b_i)$ はエントリ b_i の対象マニア名 M に対する関連度スコアである．また，m は対象期間中に対象マニア名に関する話題についてブロガー B が投稿したエントリ数である．なお，$\frac{log(m)}{m}$ では，関連性の低いエントリを大量に投稿した場合に，そのブロガーのマニア度が高くなってしまう問題に対して，エントリ数の増加の影響を緩和させている．

2.4 マニア指向ランキングの算出

本節では，マニア指向ランキングの算出方法について述べる．以下にマニア指向ランキングの算出手順を示す．

（1） ランキングの適用対象となるブログエントリ集合が与えられた際，これに対応するブロガー集合に対し，それぞれがどの話題のマニアブロガーであるかを確認し，マニアグループとして集計する．(この際，あるエントリが複数のマニアグループに属することを許す.)

（2） 上記マニアグループのうち，人数が多いものから上位 x 件（例えば 3～5 件程度）をマニア指向ランキングの対象とする．

（3） ランキング対象となったマニアグループの各ブロガーを，前節で説明したマニアスコアに基づいてソートすることで，その話題（マニアグループ）に関するランキングを行う．

以上のように検索対象であるブログエントリに対して，複数の視点からのランキングを実現する．これにより，利用者は各視点における上位ランクのブロガーが書いたエントリを選択的に閲覧することが可能になる．

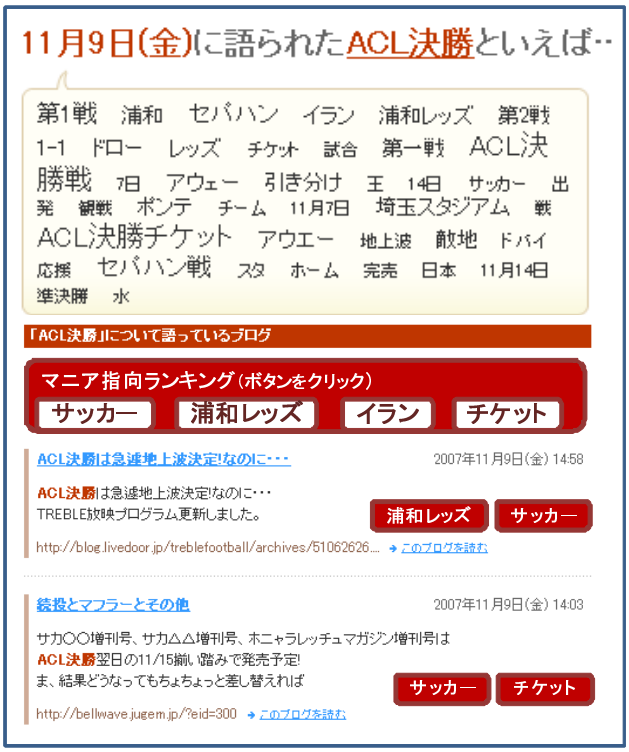

図 1　システムの実装イメージ

3. システムの実装および評価実験

3.1 システムの実装イメージ

システムの実装イメージを，図 1 に示す．図 1 は，“ ACL 決勝 ”でブログ検索された結果の例を表している．この例では，話題語が検索キーワードである“ ACL 決勝 ”であり，図の下部に“ ACL 決勝 ”にて検索されたブログエントリのリストが表示されている．従来までは，単に新着順で表示されるのみであったが，提案システムでは“ マニア指向ランキング ”を提供する．

利用者は“ サッカー ”,“ 浦和レッズ ”,“ イラン ”,“ チケット ”等のボタンをクリックすることで，その話題に関して頻繁にブログを投稿するブロガー（マニアブロガー）に基づいたランキングを表示させることができる．つまり，単に検索キーワードに基づくブログを閲覧するのではなく,“ サッカーマニアからみた ACL 決勝 ”,“ 浦和レッズマニアからみた ACL 決勝 ”,“ チケットマニアからみた ACL 決勝 ”等のように視点を選択しながらブログ情報を閲覧することが可能となる．

また，図 1 に表示されているエントリのリストには，エントリを書いたブロガーが何の話題に関するマニアブロガーかを呈示している．例えば，1 件目のブロガーは,“ 浦和レッズ ”および“ サッカー ”に関するマニアであることを示している．これにより，ブロガーの立場を把握した上でブログを閲覧することができる．

3.2 プロトタイプシステムに基づいた評価実験

前節にてシステム実装のイメージを示したが，まずは提案手法の評価を行うために評価実験用のプロトタイプシステムの構築を行った．プロトタイプシステムで扱うブログエントリ数は，2007 年 12 月 1 日以降に投稿された，13,364,604 エントリ（ブ


Result:

検索条件:1
東京(19, 9.50%),

検索条件:2
中国(24, 13.11%), サッカー(14, 11.76%), 韓国(14, 9.15%), アメリカ(18, 8.53%), テレビ(17, 6.44%), 東京(35, 5.56%), 24(13, 5.35%), 写真(20, 2.00%), 仕事(14, 1.40%),

検索条件:3
政治(21, 15.11%), 歴史(16, 13.68%), 中国(29, 10.62%), 経済(24, 10.39%), アメリカ(29, 9.86%), サッカー(17, 9.60%), 研究(21, 8.90%), 韓国(13, 5.96%), 番組(14, 5.15%), 東京(46, 4.60%), テレビ(17, 3.89%), 24(15, 3.65%), 映画(18, 2.67%), 写真(23, 2.30%), 仕事(15, 1.50%),

検索条件:4
政治(14, 11.02%), 中国(23, 9.79%), サッカー(20, 9.62%), アメリカ(18, 6.06%), 番組(17, 4.76%), 東京(40, 4.21%), テレビ(15, 3.21%), 映画(17, 2.67%), 写真(19, 1.90%), 仕事(14, 1.40%),

検索条件:5
中国(15, 14.85%), サッカー(13, 13.83%), 東京(25, 6.54%), 写真(16, 1.63%),

検索条件:6
戦争(13, 21.31%), 政治(24, 18.46%), 歴史(16, 15.38%), 中国(34, 12.36%), 経済(25, 11.52%), サッカー(20, 11.17%), アメリカ(30, 10.64%), 研究(20, 8.62%), 韓国(14, 6.86%), 番組(15, 5.66%), 東京(49, 5.33%), テレビ(20, 5.03%), 24(16, 3.98%), 大阪(13, 3.76%), 映画(17, 2.82%), 写真(22, 2.20%), 仕事(16, 1.60%),

検索条件:7
中国(27, 10.27%), サッカー(21, 10.00%), 経済(16, 7.44%), 韓国(13, 6.02%), アメリカ(18, 5.92%), 東京(42, 4.20%), 番組(18, 3.90%), テレビ(25, 3.80%), 大阪(15, 3.59%), 映画(25, 3.49%), 写真(21, 2.10%), 車(14, 1.73%),

検索条件:8
中国(28, 11.57%), サッカー(17, 8.37%), 経済(17, 8.29%), アメリカ(19, 7.76%), 韓国(14, 7.33%), 24(18, 5.56%), 東京(48, 4.80%), テレビ(21, 4.05%), 大阪(15, 4.03%), 映画(23, 3.47%), 写真(22, 2.20%),

検索条件:9
イギリス(13, 22.41%), 戦争(15, 22.06%), 歴史(22, 19.64%), 政治(24, 17.02%), ドイツ(16, 16.33%), 中国(43, 13.92%), 経済(28, 11.29%), アメリカ(32, 10.92%), サッカー(18, 9.38%), 研究(20, 9.09%), 韓国(15, 6.98%), 東京(60, 6.00%), 番組(19, 5.57%), 24(22, 5.06%), テレビ(27, 5.00%), 本(22, 4.17%), 映画(29, 3.91%), 大阪(15, 3.49%), 車(15, 2.54%), 写真(22, 2.20%), 仕事(21, 2.10%),

図 2 「オリンピック」にて検索した際のマニアランキングリスト

ロガー数：1,414,156）である（2008 年 2 月 15 日時点）．

図 2 に,「オリンピック」にて検索した際のマニアランキングリストを示す．図中の検索条件 1 ～ 9 は，2.2 節にて述べたマニアブロガーの認定条件の 1 ～ 9 に該当し，各々の条件において提示可能なマニアランキングのリストを提示している．

なお，本論文では，以下の項目に関する評価実験を行った．

- 各マニアブロガー認定条件にて提示可能なマニアランキングの項目と数の妥当性の検証．
- ユーザが選択可能なマニアランキングの項目に対するエントリの内容の妥当性の検証．
- ブロガーに対するマニアスコアの妥当性の検証．

3.2.1 マニアランキングの項目と数の妥当性の検証

ここでは，各マニアブロガー認定条件にて提示可能なマニアランキングの項目と数の妥当性の検証を行う．提示可能なマニアランキング数は安定して，ある程度の数を確保しなければならない．また，検索キーワードと全く関係ないマニアランキングの価値は高いとはいえないので，その妥当性を確保する必要がある．本節ではまず，以下の 10 個のキーワードにて検索し，平均マニアランキング数を調べた．なお，ブログエントリの検索期間は，検索時から遡って 7 日間とした．

「オリンピック」「選挙」「野球」「北京」「ラーメン」
「iPod」「Windows」「正月」「自動車」「環境」

図 3 に，各マニアブロガー認定条件における平均マニアランキング数を示す．当然ながら，検索キーワードが異なればマニアランキング名も異なっていたが，どの検索キーワードを使ってもマニアランキング数の変動は少なかった．各条件において提示可能なマニアランキングに関しては，条件 1，2，5では十分な数のマニアランキングを確保できていないといえる．条件 9 は，提示可能なマニアランキング数が最も多いが，適切とはいえないマニアランキングを多く含んでいた．マニアランキング名を確認したところ，条件 3，4，6 はある程度のマニアランキングを確保できており，また不適切なマニアランキングもそれ程多くなかった．条件 7，8 に関しては，不適切なマニアランキングを含んでいるものの，条件 3，4，6 がピックアップできていない，有効なマニアランキングをピックアップできているケースも少なからず存在した．

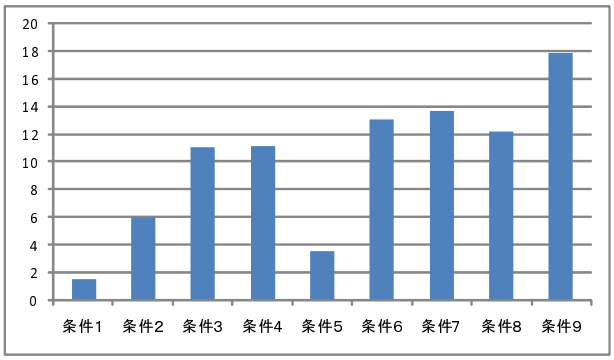

図 3 各マニアブロガー認定条件における平均マニアランキング数

したがって，3，4，6 の条件を中心に，マニアブロガーの認定条件を再検討するが，この際に条件 7，8 にてピックアップできた有効なマニアランキングをピックアップできるような条件設定を目指す．

3.2.2 マニアランキングの項目に対するエントリの内容の妥当性の検証

ある話題に対してブログ検索を行った際に，検索結果には複数の視点から書かれたブログエントリが存在するものと考えられる．提案手法では，複数のマニアランキングをユーザに提供できるため，ユーザが自分の読みたい視点から書かれたブログエントリを選択することが可能である．そこで，ある話題に対して検索した結果に対して異なるマニアランキングを選択した際に，ユーザに提示されるエントリの内容の違いに関して考察する．

例として，2007 年 12 月 3 日～10 日までに書かれたブログエントリに対して,「柔道」で検索を行った．この中で，マニア認定条件 6 により提示されるマニアランキングの上位 5 件は，以下の通りであった．

「柔道」「慶ちゃん」「ＮＥＷＳ」「ＴＶ」「ジャニーズ」

まず,「柔道」マニアランキングでは，上位 10 件のうち，3 件が有名な柔道選手に関する記事，7 件が自分で柔道をするブロガーによる記事であった．次に「慶ちゃん」マニアランキングであるが，上位 10 件全てが，アイドルグループ「NEWS」の加藤成亮さんが主演したドラマ“ 姿三四郎 ”に関する記事であった．(注：「慶ちゃん」とはジャニーズのアイドルグループ「NEWS」

	柔道	慶ちゃん	NEWS	TV	ジャニーズ
柔道		0	0	0	0
慶ちゃん	0		5	2	1
NEWS	0	5		3	3
TV	0	2	3		1
ジャニーズ	0	1	3	1	

図 4　各マニアランキング間上位 10 位以内における重複エントリ数

の小山慶一郎さんのニックネーム）「ＮＥＷＳ」マニアランキングでも，上位 10 件全てがアイドルグループ「NEWS」の加藤成亮さんが主演したドラマ“ 姿三四郎 ”に関する記事であった．ただし，上位のブログエントリが「慶ちゃん」マニアランキングと全て一致している訳ではない．「ＴＶ」マニアランキングでは，上位 10 件のうち，ドラマ“ 姿三四郎 ”に関する記事が６件，その他柔道と関係が少ないものが４件であった．「ジャニーズ」マニアランキングでは，該当ブログ記事が６件のみであったが，６件全てがドラマ“ 姿三四郎 ”に関する記事であった．

ここで，図 4 に，各マニアランキング間上位 10 位以内における重複エントリ数を示す．図 4 にて示す重複エントリは，対応するマニア間の近さを示す一つの指標であるという解釈も可能である．図 4 より，「柔道」マニアランキングは，他のランキングとの重複はなく，他のランキングに比べて特徴的であるということが言える．逆に「慶ちゃん」マニアランキングと「ＮＥＷＳ」マニアランキングでは，半数の５エントリが重複しており，マニア間の近さを証明する結果となっている．

次に，各マニアランキングにおいて，提示されるエントリのうち他のマニアランキングでは表示されないものの割合，すなわち各マニアランキング独自のエントリの割合を以下に示す．

「柔道」：100%　「慶ちゃん」：44%　「ＮＥＷＳ」：30%
「ＴＶ」：70%　「ジャニーズ」：50%

他のマニアランキングとのエントリの重複が無い「柔道」マニアの存在価値が高いことは当然であるが，「ＮＥＷＳ」マニアランキングのように他のマニアランキングとの重複エントリが数多く存在するものであっても，独自のエントリを有しており，その存在価値は決して低いとは言えない．他の類似ランキングと統合してユーザに提示することも今後検討するが，独自のエントリを利用できるような方法を検討すべきと考えている．

以上のように，マニアランキングの選択によって，提示されるブログ記事の内容が大きく異なっており，提案手法により複数のマニアランキングをユーザに提示することで，ユーザが読みたい記事を選択することが可能になるといえる．ただし，意味的に類似しているマニアランキング項目に対しては，マニアランキング同士の関連を踏まえて，ランキングの統合に関する検討や，提示方法の工夫が必要であると考えている．

3.2.3　マニアスコアの妥当性の検証

ここでは，ブロガーに対するマニアスコアの妥当性の検証として，以下に示す３つのスコアの比較を行う．

Ａ：対象キーワードを含むエントリ数のみで算出するスコア
Ｂ：式 (2) に示す提案手法のマニアスコアのうち，$\frac{log(m)}{m}$ による重み調整を行わないスコア
Ｃ：提案手法によるマニアスコア

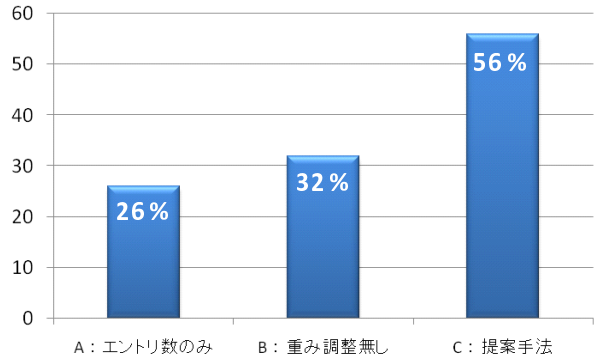

図 5　各種マニアスコアによるランキング上位 10 件に対する適合率

以下に示す５パターンの検索およびランキングに対して，上位 10 位に対するマニアスコア上位のブロガーとして妥当なものの数（適合率）を調べる．

・「キャンプ」で検索した際の「阪神」マニアランキング．
・「ニュース」で検索した際の「中国」マニアランキング．
・「テレビ」で検索した際の「嵐」マニアランキング．
・「福岡」で検索した際の「福岡」マニアランキング．
・「アメリカ」で検索した際の「経済」マニアランキング．

図 5 に，Ａ，Ｂ，Ｃにて示した各種マニアスコアによるランキング上位 10 件に対する適合率を示す．C の提案手法が最も良い結果を示した．エントリ数が異常に多いアフィリエイト目的のブログ等が，AやBではランクインしてしまう傾向がある．したがって，ノイズの排除という観点からも提案手法の有効性を確認できたといえる．しかしながら，検索およびランキングのパターンによっては手法Aが最も良いケースもあった．今後，より詳細な評価実験を行い，必要に応じてマニアスコア算出方法の改良を行うつもりである．

4.　今後の課題

・マニア候補辞書登録方法に関する検討

マニア名の候補は，Web 検索に基づいて自動で登録するが，登録されたキーワードの妥当性を検証し，ストップワードの設定を含めて，検出方法を再検討する．

今回，プロトタイプシステムでは，自動で検出できたマニア名候補は８００程度であり，十分な数を抽出できたとはいえない．また，この８００程度の語句の中にも，不適切な語句も含まれていた．すなわち，数量の面でも，精度の面でも改良の余地はあるといえる．

マニア候補名のピックアップするために利用した「マニア」「ファン」「フリーク」以外のものとしては，「大好き」「ウォッチャー」「オタク」等の語句も考えられる．また，最終的にはユーザによる登録制にすることで精度を向上させることも考えられる．したがって，あらゆる可能性を考慮しながら，マニア候補辞書登録方法の効率化を検討する．

・マニアブロガー認定基準の妥当性の検証

今回，９通りのマニアブロガー認定条件を設定して実験を行い，各条件の比較を行った．今後はさらに詳細な実験を行うことにより，より適切な認定条件について検討する．

・マニアスコアの妥当性の検証

本稿では，マニアスコアの妥当性に関して実験を行い，提案手法の有効性について議論した．検索およびランキングの条件によっては，提案手法が劣っているケースもあった．今後はより詳細な検証を行い，必要に応じて提案手法の改良を行う．

・スパム対策

大量のブログエントリを自動で投稿するようなブログサイトが存在するが，このようなサイトではマニアスコアが異常に高くなる恐れがある．意図的なスパムも含めて，これらのブログサイトに対する有効なスパム対策について検討する．

5. おわりに

本論文では，ユーザに対して複数の視点に基づくランキングの提示が可能な，ブロガーの熟知度に基づいたブログランキング方式，すなわちマニア指向ブログランキング方式を提案すると共に，実装したプロトタイプシステムに基づいた評価実験を行った．評価実験により提案手法の有効性を確認することができたが，検討すべき課題も多い．今後はさらに検証実験を行い，提案手法の改良を行う．

なお，将来的には，マニアスコアの高いブロガーに支持されるような，重要なブロガーの検出についても取り組む予定である．

謝　　辞

本研究の一部は，文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」(A01-34, 課題番号 19024058）および平成 19 年度高度通信・放送研究開発委託研究「課題ア：Web コンテンツの分析技術」（管理番号：121 ア 03）による．ここに記して謝意を表します．

文　　献

[1] S. Brin and L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In Proceedings of the 7th World-Wide Web Conference, Apr. 1998. http://www7.scu.edu.au/1921/com1921.htm.
[2] kizasi.jp, http://kizasi.jp/
[3] Yahoo!ブログ検索, http://blog-search.yahoo.co.jp/
[4] Google ブログ検索, http://blogsearch.google.co.jp/
[5] Technorati ブログ検索, http://www.technorati.jp/
[6] 竹原幹人，中島伸介，角谷和俊，田中 克己，Web 情報検索のための Blog 情報に基づくトラスト値の算出方式，日本データベース学会論文誌（DBSJ Letters）, Vol.3, No.1，pp.101-104, 2004 年 6 月.