Web検索エンジンのランキングバイアスに関する研究動向

平手　勇宇[†,††]　　吉田　泰明[††]　　山名　早人[††,†††]

† 早稲田大学メディアネットワークセンター　〒169–8050 東京都新宿区戸塚町 1–104
†† 早稲田大学理工学部　〒169–8555 東京都新宿区大久保 3–4–1
††† 国立情報学研究所　〒101–8430 東京都千代田区一ツ橋 2–1–2
E-mail: {hirate,yoshida,yamana}@yama.info.waseda.ac.jp

あらまし　Web 検索エンジンが，Web からの情報取得の一般的手段として認知されている現在，我々が Web 上から取得できる情報は，検索エンジンのランキングに依存しているとしても過言ではない．検索エンジンのランキングが公平性の欠けるランキングであった場合，我々が取得できる情報は，公平性にかけるランキングに偏った情報となってしまう．この危険性を危惧して，近年検索エンジンのランキングに関する研究が盛んに行われている．我々は，ランキングの不公平性を，ランキングバイアスと定義する．本稿では，これまで行われてきたランキングバイアスに関する研究を，ランキングバイアスを生じる原因となるインデックスカバー率，ランキングアルゴリズムに分類して論じると共に，複数の検索結果集合及びランキングの比較に関する研究について紹介する．
キーワード　検索エンジン，ランキング，バイアス，インデックス，ランキングアルゴリズム

Recent Researches in Ranking Bias of Modern Search Engines

Yu HIRATE[†,††], Yasuaki YOSHIDA[††], and Hayato YAMANA[††,†††]

† Media Network Center, Waseda University　1–104 Totsuka-cho, Shinjuku-ku, Tokyo, 169–8050, Japan
†† Science and Engineering, Waseda University　3–4–1 Okubo, Shinjuku-ku, Tokyo, 169–8555, Japan
††† National Institute of Informatics　2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 Japan
E-mail: {hirate,yoshida,yamana}@yama.info.waseda.ac.jp

Abstract　Recently, modern web search engines are recognized as fundamental tools for information retrieval from the web. It means the information that we are able to obtain may be depend on rankings of search engines. However, if search engines return unfair rankings, then following the suggestion comes up. Is the information that we obtain from the web biased according to the rankings? From this point of view, many researchers and politicians investigated about search engine rankings. We define such unfairness as ranking bias, and surveyed several current researches about ranking bias. In this paper, we report these researches by dividing two categories, comparing index coverages, analysis of ranking algorithms. Also, we report researches about comparing set of results or ranking of search engine.
Key words　Search Engine, ranking, bias, index, ranking algorithm

1. はじめに

近年，ネットワーク環境の整備によって，インターネットが日常的に利用されるようになってきている．財団法人インターネット協会のインターネット白書 2006 によると，日本の世帯の 85.4 %の世帯が，インターネットに接続していると報告されている(注1) [1]．このように，インターネットが普及するにつれて，ウェブ上に配信されている情報量も増えており，現在 Web 上には 537 億ページが存在すると推測されている [2]．このように，Web が巨大化し，膨大な情報が日々蓄積されているため，検索エンジンを利用して情報を取得する手法が不可欠になっている．

しかし，検索エンジンは，ユーザが入力したクエリに一致するページをランキングしたリストを返すシステムであり，Web 上に存在する情報を網羅した結果を返すことは保証していない．さらに，ユーザは検索結果の上位ランキングしか参照しない傾向 [3] [4] を考慮すると，ユーザが検索エンジンから得る情報は，Web 上の一部でしかない．にもかかわらず，インタース

(注1)：携帯電話等の通信機器からのインターネットアクセスも含めた値である．

コープ社の情報メディアに関する調査 [5] では,「検索エンジンによって得られたページから取得できる情報は，テレビとほぼ同等の信頼性がある」とユーザが認知していることを示している．つまり，ユーザは，検索エンジンの検索結果の上位ランクの結果から得られる限られた情報を信頼していることになる．CNET の記者である Olsen は，2002 年に,「Google によってランキングされないページは，ウェブ上に存在しないと同義である.」と言及している [6]．さらに，同じクエリを複数の検索エンジンに入力しても，検索結果に相違があることも指摘されている．2005 年に実施された Dogpile.com [7] の調査によると，Google [8], Yahoo! [9], MSN [10], Ask Jeeves [11] の 4 つの検索エンジンの上位 10 件の検索結果集合のうち，4 つの検索エンジンに共通する検索結果は，検索結果の全体の 1.1%でしかない [12](注2)．

このような状況を踏まえ，近年 CS に限らず，社会学，政治学等の研究者たちの間で，検索エンジンのランキングの公平性に関する議論や研究が行われ始めた．ランキングの公平性とは，ウェブ上に存在するページが，記述されている内容のクオリティに従って正しくランキングしているか否かを指す．つまり，公平性のあるランキングとは，上位ランクに信頼性のある情報が掲載されているクオリティの高いページが並んでいるランキングであり，理想的なランキングである．逆に信頼性のないクオリティの低いページが上位ランクにランキングされる場合，そのランキングは不公平なランキングとなり，理想的でないランキングとなる．以上を踏まえると，すべての検索エンジンのランキングが公平性のある理想的なランキングであったとするならば，すべての検索エンジンは，同じキーワードに対して同じ検索結果ランキングを返すことになる．しかし，ページのクオリティを図ることは困難であり，現在の検索エンジンでは，同じキーワードでも，違った検索結果ランキングを返している．したがって我々は，ランキングの不公平さ，すなわち理想的なランキングからの “ズレ” を，検索エンジンのランキングバイアスとして定義する．本稿では，CS の観点から見たランキングバイアスに関連する近年の研究を調査し，ランキングバイアスの研究動向を報告する．

本稿では，次のような構成をとる．第 2 節では，近年の研究で問題視されている，“rich-get-richer” 現象，“Googlearchy” について説明する．第 3 節では，検索エンジンのバイアスを引き起こす要因に関して述べる．第 4 節では，近年のランキングバイアスに関連する研究傾向を述べる．その後，近年の検索エンジンのバイアスに関連する研究を，インデックスカバー比較，ランキングアルゴリズムの検討の 2 つのカテゴリに分類し，それぞれ第 5 節，第 6 節で紹介する．加えて，ランキングの直接比較の研究を第 7 節で紹介する．最後に第 8 節でまとめを行う．

(注2)：同社の検索エンジンに蓄積されたアクセスログから 12,570 のクエリを抽出し，実際に検索を実行して計測している．

2. 検索エンジンで問題視されている現象

本節では，ランキングバイアスを考える上で，現在の検索エンジンで問題視されている現象について述べる．

2.1 “rich-get-richer” 現象 [13]

“rich-get-richer” 現象とは，既に検索エンジンのランキングのトップにランクされているウェブページは，ユーザの注目を得やすく，さらに人気度を高めやすい現象を指す [13]．この現象の裏には，現在ランキングの下位にランクされているページ，または新規に生成されたページは，人々の注目を得ることが困難で，人気度を高められない傾向があることを意味する．

たとえば，次に示すようなケースが考えられる．
ある授業で「靖国参拝」に関するレポートが出されたとする．学生は検索エンジンを用いて「靖国参拝」とクエリを指定して検索をかける．レポート提出期限の制約で，学生は検索結果の Top10 にランキングされているページを参照してレポートを書くとする．この検索結果の Top10 には，靖国参拝を反対するサイトがランクインして，Top10 以下に靖国参拝に賛成するサイトがランクインしたとする．この時，学生は靖国参拝を反対する意見に沿ったレポートを作成して，そのレポートをウェブ上に公開する．作成したページは，おそらくは Top10 にランクインしたサイトにリンクが張られている．

つまり，この一連の行動で，Top10 にランクされたサイトはさらにインリンクを集め，人気度を高める．しかし，Top10 にランクされなかったサイトは，インリンクを集めることができず，人気度を高めることができない．

この現象が問題視されている理由は，どんなにクオリティの高いコンテンツを書いたとしても，それが検索エンジンによってランキングされることが困難であるからである．さらに，当該コンテンツがランキングされるまで，長い時間がかかるという問題がある．その結果，良質のコンテンツがなかなか上位にランキングされないという観点から，検索エンジンの検索結果の質を落とす可能性があることが示唆されている [14].

2.2 “rich-get-richer” 現象の検証

UCLA 大学の Cho らは文献 [13] で，実際の Web ページを用いて rich-get-richer 現象を検証している．検証対象の Web ページは，Open Directory [15] に登録されている 154 個のホストをランダムに選択し，選択した 154 ホストを 7ヶ月の間隔を設けて 2 度フォーカストクローリングして得られた 2 つのページセットである．検証方法は以下のとおりである．

（1） 1 回目に取得したページセットをもとに，全てのページのインリンク数，及び PageRank [16] を算出する．

（2） インリンク数の降順にページを並び替え，取得した全てのページを上位から順番に 10 個のグループに分割する．10 個のグループは，同数のページで構成される．

（3） それぞれのグループごとに，グループに属する全てのページのインリンク数合計値と，PageRank 合計値を計算する．

（4） 2 回目に取得したページセットをもとに，全てのページのインリンク数と PageRank を算出し,（2）で生成したグ

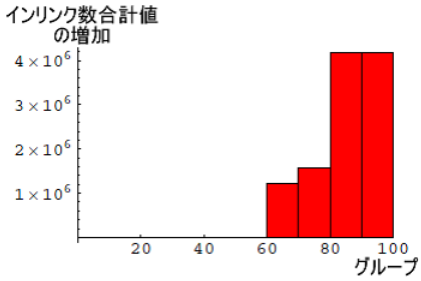

図 1　インリンク数によるグループと，インリンク数合計値の増加 ([13] より引用)

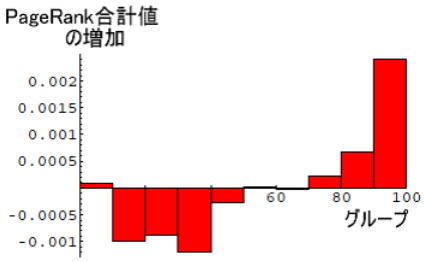

図 2　インリンク数によるグループと，PageRank 合計値の増加 ([13] より引用)

ループごとに，インリンク数合計値，PageRank 合計値を計算する．

（5） 全てのグループに対して，1 回目のページセットのインリンク数合計値と，2 回目のページセットのインリンク数合計値の差を計算する．同様に，PageRank 合計値の差も計算をする．

以上の方法でそれぞれのグループで，インリンク数，PageRank の増減を計算し，rich-get-richer 現象を検証している．グループごとのインリンク数合計値の増減のヒストグラムを図 1 に，PageRank 合計値の増減のヒストグラムを図 2 に示す．図 1，図 2 より，1 度目の収集時に既に多くのインリンクを得ていたページが，2 度目の収集時までに，さらに多くのインリンクを得ており（図 1），PageRank も増加させていることがわかる（図 2）．これにより，rich-get-richer 現象が確認できる．

2.3　“Googlearchy” [17] [18]

前述の rich-and-richer 現象の結果，近年の Web は，膨大なインリンクを持ったごく少数のページ集合と，インリンクをほとんど持たない多数のページ集合で構成されていると考えることができる [17] [19] [20]．そして，検索エンジンは，膨大なインリンクを持ったごく少数のページを検索結果のトップにランキングするので，ユーザのトラフィック（ページ訪問）は，これらのページに集中する．このランキングによるトラフィック集中の現象を，Google のランキングが加速している想定して，“Googlearchy” と呼ばれるようになった [17] [18]．逆にランダムウォークモデル [13] と比較して，トラフィック集中が発生していない場合を”Googlocracy”と呼んでいる [18]．

3.　ランキングバイアスを引き起こす要因

第 1 節で示したように，本稿では，ウェブ上に存在するページが，記述されている内容のクオリティに従ってランキングされているものを，公平性のあるランキングと定義する．ページのクオリティ順に並んでないランキングを，公平性にかけるランキングとして，ランキングバイアスと定義する．たとえば，非常にクオリティが高いページが，ランキングの下位に出現したり，ランキングには出現しなかった場合は，ランキングバイアスである．同じように，クオリティの低いページが，ランキングの上位に出現した場合も，ランキングバイアスである．検索エンジンのランキングバイアスを引きこす要因は多数存在する．本節では，ランキングバイアスの引き起こす要因について述べる．

3.1　インデックスの網羅性

検索エンジンは，ウェブをクローリングすることで検索エンジンのインデックスを構築している．そして，インデックスに登録されていないウェブページは，検索エンジンには含まれない．検索エンジンのインデックスに登録されていないウェブページの中で，非常に高いクオリティのページが存在し，検索結果に含まれない場合，検索エンジンのランキング結果は，公平性に欠け，バイアスのかかった検索結果となる．

3.2　ランキングアルゴリズム

現在の多くの検索エンジンは，PageRank [16] アルゴリズムや，PageRank に類するアルゴリズムにより人気度を間接計算している．PageRank の基本は，“コンテンツのクオリティが良いページは，良質の多くのページからリンクされている．”であり，コンテンツのクオリティとインリンク数が比例関係にあることを前提条件としている．

しかし，2.1 で述べた “rich-and-richer” 現象が存在する近年の Web 上においては，コンテンツのクオリティとインリンク数の比例関係が崩れている可能性がある．したがって，PageRank や，PageRank に類するアルゴリズムを用いている検索エンジンのランキングは，バイアスのかかった検索結果となる可能性がある．

3.3　意図的なランキング

意図的なランキングとは，人の手によって故意に改変されたランキングのことをさす．例として，特定の Web サイトが検索結果に表示させないように設定することや，特定の Web サイトのランキングを本来あるべきランキングよりも高い，もしくは低いランキングに設定することが考えられる．検索結果に，このような意図的なランキングが含まれていた場合，その検索結果はバイアスのかかったものとなる．

4.　ランキングバイアスに関する研究動向

第 3 節で示したように，検索エンジンのバイアスを引き起こす原因は複数存在する．ランキングバイアスを引き起こす要因を調査するために，近年，さまざまな研究が行われている．具体的には，3.1 で示したインデックスの網羅性を調査するために，インデックスカバー比較に関する研究 [21] [22] [23] [24] や，

3.2 で示したランキングアルゴリズムの妥当性を調査するための研究 [13] [14] [31] [32]，さらには，検索エンジンの検索結果集合またはランキングの比較を試みた研究 [35] [36] [37] がある(注3)．

本稿では，インデックスカバー率に関する研究を第 5 節で，ランキングアルゴリズムの妥当性に関する研究を第 6 節で，検索結果集合及びランキングの比較に関する研究を第 7 節で述べる．

5. インデックスカバー比較に関する研究

3.1 で述べたとおり，検索エンジンのインデックスの網羅性によって，ランキングのバイアスが発生する可能性がある．本節では，検索エンジンのインデックス網羅性を比較した研究について述べる．

検索エンジンのインデックス網羅性比較の調査研究は，98 年の Bharat らによる研究 [21]，99 年の Henzinger らによる研究 [22]，2004 年の Vaughan らによる研究 [23]，2006 年の Yossef らによる研究 [24] が存在する．

DEC の Bharat(注4) らは，97 年に文献 [21] で，クエリ集合のサンプルを検索エンジンに取り合わせることによって，AltaVista [25]，Excite [26]，HotBot [27]，Infoseek(注5) [28] のインデックスサイズを比較している [21]．

Compaq の Henzinger(注6) らは，98 年に文献 [22] で，Web 上をランダムウォークすることで得られたページ集合を検索エンジンに問い合わせることで，検索エンジンのインデックスを比較している．調査対象の検索エンジンは，AltaVista [25]，HotBot [27]，Excite [26]，Infoseek [28]，Google [8]，Lycos [29] である．

Western Ontario 大学の Vaughan らは，文献 [23] で，2004 年にアメリカ，中国，台湾，シンガポールの 4 カ国の商用サイトのインデックス網羅率を調査している．調査対象の検索エンジンは，Google [8]，AltaVista [25]，AlltheWeb [30] の 3 つである．網羅率調査方法は以下のとおりである．

（1） アメリカからは 143 ホスト，中国からは 143 ホスト，台湾からは 141 ホスト，シンガポールからは 94 ホスト選択する．

（2） それぞれホスト内の Web ページを対象にフォーカストクローリングによってウェブページをすべて取得する．

（3） フォーカストクローリングによって得られたページ数を 100 %として，3 つの検索エンジンの当該ホストのインデックス数と比較する(注7)．

結果は図 3 に示すとおりであった．図 3 は，Google がすべての国において他の 2 つのインデックスよりも高い網羅率を保持

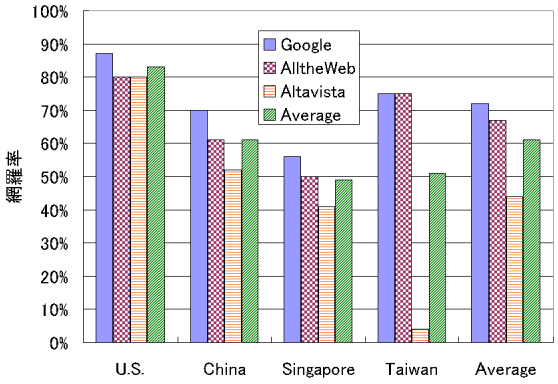

図 3　4 カ国の商用サイトの網羅率 ([23] のデータを基に生成)

表 1　Google, MSN, Yahoo!のインデックス相対比較（ [24] より引用）

Page from ↓ Indexed by →	Google	MSN	Yahoo!
Google		46%	45%
MSN	55%		51%
Yahoo!	44%	22%	

図 4　Google, MSN,Yahoo!のインデックス相対比較（ [24] のデータより作成）

しているのに対し，AltaVista は台湾の商用サイトのインデックス網羅率が極端に低い値であることを示している．

Technion（イスラエル工科大学）の Bar-Yossef らは，2006 年に文献 [24] で，検索エンジンのインデックスのサンプリング手法を提案し，Google [8]，Yahoo! [9]，MSN [10] の相対的なインデックス比較を行っている．インデックスの相対比較結果（2006 年）は，表 1 のとおりである．ただし, [24] の提案手法は，英単語の Query-Pool を生成し，Query-Pool をサンプリングして検索エンジンにクエリを投げる手法である．それぞれの検索エンジンから返ってきた URL 集合の包含関係を計算することで，インデックスの相対比較を行っている．したがって，表 1 の結果は 3 社の英語ページのインデックス比較であることに注意が必要である．表 1 をベースに 3 つの検索エンジンの英語ページのインデックスカバーの相対比較を図示したものが図 4 となる．表 1 と図 4 より，英語ページをもっとも多くインデックスしているのは，Yahoo!であるという結果であった [24]．

6. ランキングアルゴリズムに関する研究

3.2 で述べたとおり，PageRank アルゴリズムや，その類似アルゴリズムが，公平性のあるランキングアルゴリズムであるか否かを調査する研究がおこなわれている [13] [14] [31] [32]．このカテゴリに属する研究は，ユーザの訪問 V，ランキング R，人気度 P のモデル化を行って検討している．

(注3)：3.3 で述べた意図的なランキングを主対象とした調査研究は，現時点で存在していない．
(注4)：現在は Google に勤務
(注5)：現在は Go.com となり Yahoo!の検索エンジンを利用している．日本では，楽天が運営してる．
(注6)：現在は Google に勤務
(注7)：Google では”site:ホスト名”，AltaVista では”host:ホスト名”，AlltheWeb では”ホスト名”をクエリに指定することで，ホストのインデックス数を調べることができる．

文献 [13] [31] では，検索エンジンをまったく利用しないユーザ行動をモデル化した**ランダムサファーモデル**と，検索エンジンのみを利用したユーザ行動をモデル化した**検索エンジン支配モデル**を構築し，両者の比較を行っている．文献 [14] では，公平性のあるランキングを目指して，ページのクオリティをリンク構造によりモデル化を行う手法を提案している．文献 [32] では，検索結果の上位に，ランキングとは関係のないランダムなページを挿入することで，2.1 で示した rich-get-richer 現象が，どの程度緩和されるかを，モデルを用いて検証している．

本節では，ランダムサファーモデルと検索エンジン支配モデルの比較を 6.1 で，ページクオリティのモデル化を 6.2 で，rich-get-richer 現象を緩和するためのランキングにランダム要素を挿入手法について 6.3 で述べる．

6.1 ランダムサファーモデルと検索エンジン支配モデルの比較 [13] [31]

UCLA の Cho らは，文献 [13] において，ランダムサファーモデルと検索エンジン支配モデルの比較を行い，新規ページが人気度を得るまでの時間の比較を行っている．その結果，新規ページが人気度を得るまでの時間は，検索エンジン支配モデルのほうが，ランダムサファーモデルよりも 60 倍の時間を要する結果を算出している．また，文献 [31] では，実データを用いて，実際のユーザ行動はランダムサファーモデルと検索エンジン支配モデルのどちらに近いかを検証している．

6.1.1 2つのモデルの定義

ランダムサファーモデルとは，検索エンジンのランキング結果を一切使用しないユーザ行動モデルで，次に示す 2 つの前提条件を満たしているものとする．

（1） 時間 t におけるあるページ p の訪問数 $V(p,t)$ は，時間 t におけるあるページ p の人気度 $P(p,t)$ に比例する．

（2） 全てのページは，全てのユーザから等しい確率で訪問される．

これに対し，検索エンジン支配モデルは，検索エンジンのランキングを使用してのみページ訪問をすることを指す．検索エンジンのランキングから得られるウェブページに含まれるリンクはたどらないこととする．

6.1.2 新規ページが人気度を得られるまでの時間の比較 [13]

ランダムサファーモデルでは，前述した前提条件 (1) より，時間 t における訪問数と人気度には比例関係が成立するので，以下の式が成立する．

$$V(p,t) \sim P(p,t) \tag{1}$$

さらに，$P(p,t)$ の時間 t による変化をモデル化すると図 5 のようになる(注8)．図 5 より，$P(p,t)$ は 3 つの時期に分割される．

幼少次期 $(0 \leqq t < 13)$ ユーザにほとんど訪問されることもなく，人気度も上昇しない次期

拡大時期 $(13 \leqq t < 25)$ 訪問が繰り返されることによって，人気度が一気に上昇する次期

(注8)：詳しい証明は [14] を参照のこと．また，このモデルにおいて，ページのクオリティ $Q(p)(0 \leqq Q(p) \leqq 1)$ は 1 とした．

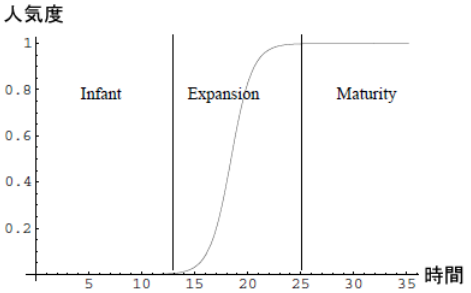

図 5 時間経過によるランダムサファーモデルの人気度の変化 ([14] より引用)

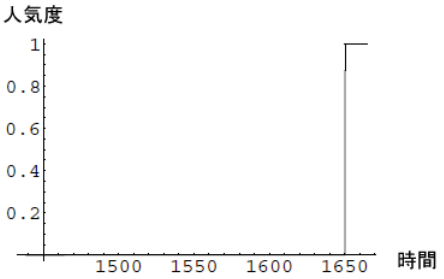

図 6 時間経過による検索エンジンモデルの人気度の変化 ([14] より引用)

成熟時期 $(t \geqq 25)$ ページクオリティ $Q(p)$ に見合った人気度に収束する次期

これに対し，検索エンジン支配モデルでは，訪問数と人気度の関係は，以下の式が成立する．

$$V(p,t) \sim P(p,t)^{2.25} \tag{2}$$

これは，文献 [4] において，クエリログのクリックスルーデータを用いて解析した結果，訪問数 $V(p,t)$ とランキング $R(p,t)$ には，

$$V(p,t) \sim R(p,t)^{-1.5} \tag{3}$$

の関係が成立すると報告されており，さらに，Stanford WebBase プロジェクト [33] [34] のデータを用いて，ランキング $R(p,t)$ と人気度 $P(p,t)$ には，

$$R(p,t) \sim P(p,t)^{-1.5} \tag{4}$$

が成立するという実験結果をもとに算出されたものである．

同じように，ランダムサファーモデルと同じ条件で，検索エンジン支配モデルにおける $P(p,t)$ をモデル化すると，図 6 のようになる．図 6 では，拡大時期に到達するまでに時間 1650 が必要であり，ランダムサファーモデルと比べて 60 倍もの時間が必要であると結論づけている [13]．

6.1.3 実データと 2 つのモデルの比較 [31]

Indiana 大学の Fortunato らは，2006 年に文献 [31] において，2 つのモデルと実データを比較し，ユーザの行動モデルは 2 つのモデルのうちどちらに近いのかを検証している．文献 [13] では，$V(p,t) \sim P(p,t)^{2:25}$ が成立していると主張しているのに

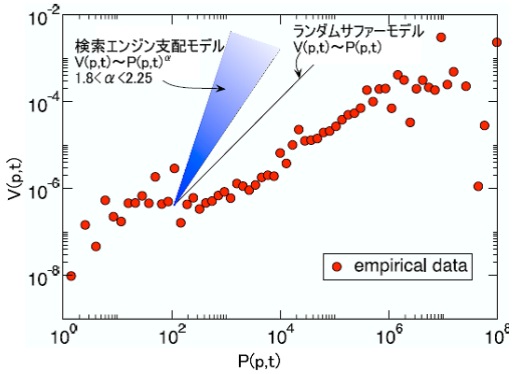

図 7　実データを基にした人気度と訪問数の関係 ([31] より引用)

対し，文献 [31] では，$V(p,t) \sim P(p,t)^{1.8}$ の関係が成立すると主張している．実データの $V(p,t)$ と $P(p,t)$ の関係をプロットしたものが図 7 である．図 7 より，文献 [31] では，実際のデータは, [13] [31] の検索エンジン支配モデルのどちらにも似ていなく，ランダムサファーモデル $V(p,t) \sim P(p,t)$ に近いとしている．よって，現状のユーザのウェブ上の行動は，Googlearchy ではなく，Googlocracy であると主張している．

6.2　ページのクオリティのモデル化 [14]

UCLA の Cho らは，文献 [14] において，ページのクオリティをウェブページのリンク情報をもとにモデル化を行い，クオリティの計算方法を提案している．PageRank [16] は，重みつきのインリンク数によって，当該ページのクオリティに比例関係があるという考えのもと，重みつきインリンク数の合計を当該ページのクオリティと見立てていた．しかし，3.2 で述べたように，新規に生成されたページや，現在低いランクに位置づけられているページは，PageRank の考えは必ずしも当てはまらない．

そこで文献 [14] では，公平性のある $Q(p)$ によるウェブページのランキングを目的として，ページ p のクオリティ $Q(p)$ を，人気度 $P(p,t)$ の式によってモデル化を行い検証している．

6.3　ランキングへのランダム要素挿入 [32]

CMU の Pandey らは，文献 [32] において，新規生成されたクオリティの高いページの人気度を上げることを目的として，新規ページの上位ランクインのための機会を提供するメカニズムを提案している．このメカニズムは，検索エンジンの上位ランクにランダム要素を挿入することで達成できるとし，Randomized Rank Promotion と呼ばれている．

Randomized Rank Promotion をモデル化することで検証を行った結果，図 5 で示した時間経過による人気度 $P(p,t)$ の変化の幼少時期を短縮できることが報告されている [32].

7.　検索結果集合・ランキングの比較に関する研究

本節では，実際に検索エンジンの検索結果集合・ランキングを比較する研究 [35] [36] [37] を紹介する．文献 [35] [36] では，複数の検索エンジンの検索結果をもとに，個々の検索エンジンのバイアス値を定義し，検索結果の偏りを定量的に表現している．また，文献 [37] では，同一クエリを同一検索エンジンに投げることで，時間経過に伴う検索結果のランキングの比較を行っている．

7.1　ランキングバイアスの計測 [35] [36]

ニューヨーク州立大学の Mowshowitz らは，検索結果の偏りを定量的に表現するために，個々の検索エンジンのバイアス値(注9)を定義している [35] [36]．ただし, [35] [36] では，検索結果のランキングは考慮していない．検索結果の上位ランクの要素と下位ランクの要素を同等に扱っていることを併記しておく．

文献 [35] [36] において，理想的な検索結果は，複数の検索エンジンの結果を結合したものと近似していると考えている．具体的には，ある検索エンジン E のバイアス $B(E)$ は，E によって検索される結果のセットで表される検索結果ベクトルと，複数の検索エンジン集合 C によって検索される結果セットとの類似度 $Sim(E)$ を用いて，

$$B(E) = 1 - Sim(E) \tag{5}$$

と表現している．

7.1.1　バイアスの定量的な表現

たとえば，3 つの検索エンジン E_1, E_2, E_3 に対し，クエリ q_1 を検索した結果，E_1 が URL リスト (u_1, u_2, u_3)，E_2 が URL リスト (u_4, u_3, u_5)，E_3 が URL リスト (u_3, u_1, u_4) を返したとする．この場合，各検索エンジンの検索結果ベクトル S はそれぞれ

$$S_1 = (1,1,1,0,0),\ S_2 = (0,0,1,1,1),\ S_3 = (1,0,1,1,0)$$

と表現される．理想的な検索結果ベクトル S は，全ての検索エンジンの検索結果ベクトルの同じ項を足すことによって生成する．この場合は，

$$S = (2,1,3,2,1) \tag{6}$$

となる．類似度 $Sim(E_i)$ は，S とそれぞれの S_i のコサイン尺度で計算される．その結果，検索エンジン E_1, E_2, E_3 のバイアス $B(E_1), B(E_2), B(E_3)$ は，以下のとおりになる．

$$B(E_1) = 1 - \frac{2+1+3}{\{3\cdot(4+1+9+4+1)\}^{-0.5}} = 0.205 \tag{7}$$

$$B(E_2) = 1 - \frac{3+2+1}{\{3\cdot(4+1+9+4+1)\}^{-0.5}} = 0.205 \tag{8}$$

$$B(E_3) = 1 - \frac{2+3+2}{\{3\cdot(4+1+9+4+1)\}^{-0.5}} = 0.073 \tag{9}$$

7.1.2　実　　験

文献 [36] では，実際に About, Ah-ha(注10), AltaVista [25], AOLSearch, FastSearch, Google [8], LookSmart, Lycos [29], MSN [10], Netscape, Overture, Sprinks, Teoma, TureSearch, WiseNut, Yahoo! [9] の 16 検索エンジンを用いて，バイアスの計算を行っている．使用するクエリは, [38] [39] から，5 カテゴリに分かれる 25 クエリをそれぞれ生成している. [38] から生成した 16 検索エンジンの 25 個クエリ分のバイアス値のプロット

(注9)：次文で述べるように, [35] [36] では，ページのランキングを考慮していないので，ここでは，検索エンジンのバイアスとした．
(注10)：現在の Enhance Interactive

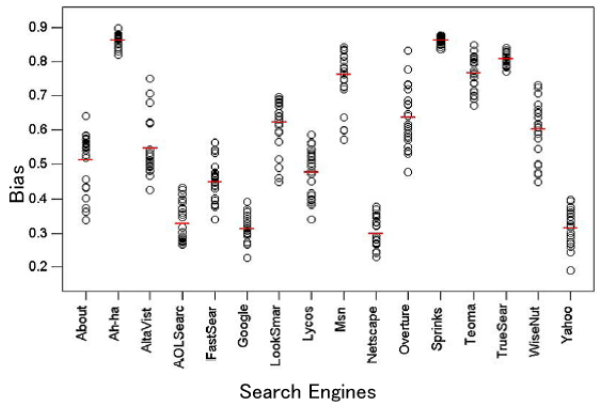

図 8 [38] から生成したクエリを入力した場合の各サーチエンジンのバイアス値（[36] より引用）

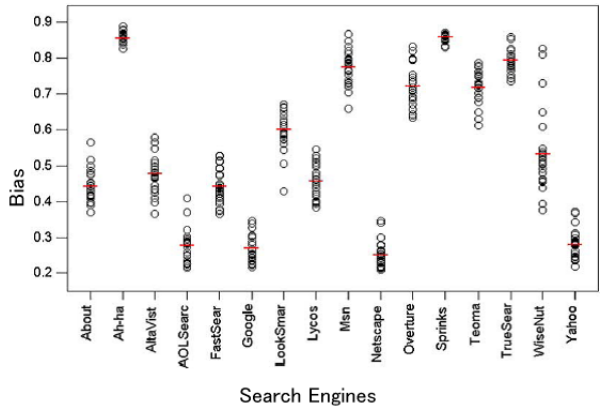

図 9 [39] から生成したクエリを入力した場合の各サーチエンジンのバイアス値（[36] より引用）

を図 8 に, [39] から生成した 16 検索エンジンの 25 個クエリ分のバイアス値のプロットを図 9 に示す.

図 8，図 9 の両者ともに AOL Search, Google, Netscape, Yahoo!が低いバイアス値を示しており，Ah-ha，MSN, Sprinks, Teoma, TrueSearch が高いバイアス値を示していた [36].

この 16 の検索エンジンは，それぞれが構築しているインデックスサイズに大きな差があると考えられる．我々は，このバイアス値の違いは，主に，3.1 で示したインデックスの網羅性の違いによって引き起こされているものと解釈している．

7.2 時間変化に伴うランキング変化の計測

Bar-llan 大学の Bar-llan らは 2006 年に，同一検索エンジンに 21 日間毎日同じクエリを投げることで，検索結果のランキングの時間経過に伴う変化を調査している [37]．文献 [37] では，テキスト検索として Google, Yahoo!, Teoma(注11) に対して 3 つのクエリ(注12)を，イメージ検索として Google Image, Yahoo! Image, Picsearch [40] に対して 2 つのクエリ(注13)を選定している．また，ランキングの比較には，共通検索結果集合数，Spearman's footrule [41] [42]，Fagin's measure [43]，そして新たな指標である M measure [37] の 4 つの指標を用いて定

(注11)：Teoma は，現在 Ask.com [11] に吸収されている．
(注12)：“US elections 2004”，“organic food”，“DNA evidence”
(注13)：“Twin towers”，“Bondi beach”

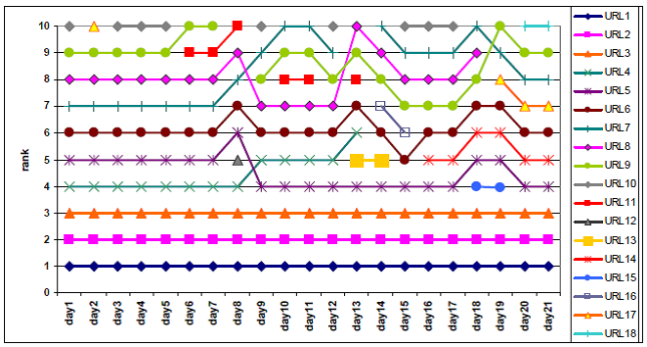

図 10　キーワード “DNA evidence” における Google のテキスト検索上位 10 件の検索結果の時間変化（[37] より引用）

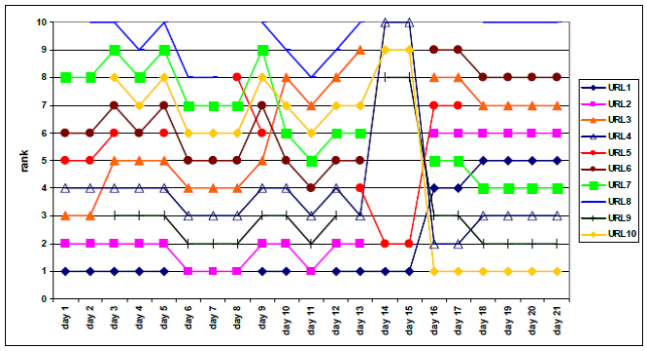

図 11　キーワード “Bondi beach” における Yahoo!のイメージ検索上位 10 件の検索結果の時間変化（[37] より引用）

量的に表現しているが，本稿では紙面の都合上割愛する．

7.2.1 実験結果

図 10 に，テキスト検索の結果として，検索エンジン Google, 使用キーワード “DNA evidence” とした時のランキング変化を示す．また，図 11 に，イメージ検索の結果として，検索エンジン Yahoo!，使用キーワード “Bondi beach” とした時のランキング変化を示す.

図 10 に示すように，Google のテキスト検索エンジンは，キーワード “DNA evidence” において検索結果上位 10 件のランキングは 21 日間でほとんど変化していないことがわかる．この傾向は，すべてのキーワード，すべての検索エンジンについても同様のことが言えると，文献 [37] では報告している．

また，図 11 に示すように，Yahoo!のイメージ検索は，キーワード “Bondi beach” において，検索結果上位 10 件のランキングは，テキスト検索に比べて変化が大きいことがわかる．Google についても，テキスト検索よりも，イメージ検索の検索結果のほうが，ランキングが激しい結果が報告されている(注14)

8. おわりに

本稿では，近年問題視されている検索エンジンのバイアスに関しての研究動向を調査した．検索エンジンのバイアスを発生

(注14)：上述した 4 つの指標の数値は，いずれも Google イメージ検索よりも，Yahoo!イメージ検索のほうが，ランキングの変化が大きい数値であった [37]．また，Picsearch [40] は，ランキングの変化はほとんどない結果を示す数値であった．

させる要因として，インデックスの網羅率，ランキングアルゴリズム，意図的なランキングの 3 つがあることを述べた．

インデックスの網羅率については，1998 年から検索エンジン間のインデックスの網羅率の比較研究が行われてきた．しかし，検索エンジンのインデックスは，検索エンジンが提供する API や公的なインターフェースを用いてしか調査できないため，サンプリングの手法を用いて推定するしかない．このサンプリングの手法の精度によって，インデックスの網羅率の精度は変わるため，今後，さらなるサンプリング手法が必要とされるものと考えられる．

また検索エンジンのランキングアルゴリズムを問題視し，モデル化によって検討した研究は 2004 年から出現し始めたばかりである．モデル化の研究では，使用されているデータの規模が小さく，その精度が問題である．つまり，大規模なデータを用いて検証した場合に違った結果が出てくる可能性があることに留意しなければならない．

さらに，検索エンジンの検索結果ランキングを直接比較する研究では，検索結果集合をランキングの違いによらず一様に扱っていること，あるいは比較するランキング対象が上位 10 件と上位ランキング鑿を扱っていることが改善点として考えられる．

我々は，こうした検索エンジンバイアス関連研究のサーベイをもとに 2006 年 10 月から主要な検索エンジンのランキング調査を継続的に実施しており，今後，その結果について報告予定である．

文　献

[1] 財団法人インターネット協会, “インターネット白書 2006,” インプレス R & D, 2006.
[2] Y. Hirate, S. Kato, and H. Yamana, “Web Structure in 2005.” Proc. of WAW2006, 2006.
[3] T. Josvhims, “Optimizing Search Engine using Clickthrough Data,” Proc. of ACM SIGKDD2002, pp. 133–142, 2002.
[4] R. Lempel and S. Moran, “Predictive Caching and Prefetching of Query Results in Search Engines,” Proc. of WWW2003, pp. 19–28, 2003.
[5] 株式会社インタースコープ自主公開レポート, “情報メディアに関する調査,” 2006.
[6] S. Olsen, “Does search engine’s power threaten Web’s independence?,” http://news.com.com/2009-1023-963618.html, 2002.
[7] Dogpile Web Search Home Page, http://www.dogpile.com
[8] Google, http://www.google.com/
[9] Yahoo!, http://www.yahoo.com/
[10] MSN, http://www.msn.com/
[11] Ask.com Search Engine, http://www.ask.com/
[12] Dogpile.com, “Different Engines, Different Results,” http://compareseachengines.dogpile.com/OverlapAnalysis.pdf
[13] J. Cho and S. Roy, “Impact of Search Engines on Page Popularity,” Proc. of WWW2004, pp. 20–29, 2004.
[14] J. Cho, S. Roy, and R. E. Adams, “Page Quality: In Search of an Unbiased Web Ranking,” Proc. of ACM SIGMOD2005, pp. 551–562, 2005.
[15] Open Directory, http://dmoz.org
[16] L. Page, S. Brin, R. Motwani, and T. Winograd, “The Pagerank Citation Ranking: Bringing Order to the Web,” Technical Report, Stanford University, 1998.
[17] M. Hindman, K. Tsioutsiouliklis, and J. A. Johnson, “Googlearchy: How a Few Heavily-Linked Sites Dominate Politics on the Web,” Annual Meeting of the Midwest Political Science Association, 2003.
[18] F. Menczer, S. Fortunato, A. Flammini, and A. Vespignani, “Googlearchy or Googlocracy?,” IEEE Spectrum, 2006.
[19] R. Albert, A.-L. Barabasi, and H. Jeong, “Diameter of the World Wide Web,” Nature, Vol. 401(6749), pp. 130–131, 1999.
[20] A. Broder, R.Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, “Graph Structure in the Web,” The International Journal of Computer and Telecommunication Networking, Vol. 33, pp. 309–320, 2000.
[21] K. Bharat and A. Broder, “A Technique for Measuring the Relative Size and Overlap of Public Web Search Engines,” Computer Networks and ISDN Systems, Vol. 30, Issue 1–7, pp. 379–388, 1998.
[22] M. Henzinger, A. Heydon, M. Mitzenmacher, and M. Najork, “Measuring Index Quality using Random Walks on the Web,” Proc. of WWW1999, pp. 213–225, 1999.
[23] L. Vaughan and M. Thelwall, “Search Engine Coverage Bias: Evidence and Possible Causes,” Information Processing and Management, Vol. 40, No. 4, pp. 693–707, 2004.
[24] Z. Bar-Yossef and M. Gurevich, “Random Sampling from a Search Engine’s Index,” Proc. of WWW2006, pp. 367–376, 2006.
[25] AltaVista, http://www.altavista.com/
[26] Excite, http://www.excite.com/
[27] HotBot, http://www.hotbot.com/
[28] Go.com, http://www.go.com/
[29] Lycos, http://www.lycos.com/
[30] AlltheWeb, http://www.alltheweb.com/
[31] S. Fortunato, A. Flammini, F. Menczer, and A. Vespignani, “The Egalitarian Effect of Search Engine,” eprint arXive:cs.CY/0511005, 2006.
[32] S. Pandey, S. Roy, C. Olston, J. Cho, and S. Chakrabarti, “Shuffling a Stacked Deck: The Case for Partially Randomized Ranking of Search Engine Results,” Proc. of VLDB2005, pp. 781–792, 2005.
[33] H. Hirai, S. Raghavan, H. G-Molina, and A. Paepcke, “Webbase: A repository of the Web,” Proc. of WWW2000, pp. 277–293, 2000.
[34] The Stanford WebBase Project, http://dbpubs.stanford.edu:8091/testbed/doc2/WebBase/
[35] A. Mowshowitz and A. Kawaguchi, “Bias on the Web,” Communications of the ACM, Vol. 45, No. 9, pp. 56–60, 2002.
[36] A. Mowshowiz and A. Kawaguchi, “Measuring Search Engine Bias,” Information Processing and Management, Vol. 41, pp. 1193–1205, 2005.
[37] J. Bar-llan, M. Mat-Hassan and M. Levene, “Methods for Comparing Rankings of Search Engine Results,” eprint arXive:cs.IR/050539, 2006.
[38] West Publishing, “West’s analysis of American low,” Eagan, MN: Thomson-West, 2002.
[39] Library of Congress Classification System, http://www.loc.gov/
[40] Picsearch - Image Search for Pictures and Images, http://www.picsearch.com/
[41] P. Diaconis, R. L. Graham, “Spearman’s Footrule as a Measure of Disarray,” Journal of the Royal Statistical Society, Series B(Methodological), Vol. 39, pp. 262–268, 1977.
[42] C. Dwork, R. Kumar, M. Naor and D. Sivakumar, “Rank Aggregation Methods for the Web,” Proc. of WWW2001, pp. 613-622, 2001.
[43] R. Fagin, R. Kumar and D. Sivakumar, “Comparing Top k Lists,” SIAM Journal on Discrete Mathematics, Vol. 17, No. 1, pp. 134–160, 2003.