5Q-04

WWW における検索ランキングとその Social Filtering としての効果*

三浦 信幸† 高橋 克巳‡ 島 健一§

NTT ソフトウェア研究所

E-mail: {miura , takahasi, kshima}@slab.ntt.co.jp

1 はじめに

インターネット・タウンページ [1] では，今年の 5 月下旬より，ユーザの検索ランキングを「インターネットタウンページ *Best HITS!*」[1] として公開している．本稿では，これを題材として，Social Filtering[2, 3] の最も簡便な形式としての検索ランキングの効果を分析する．

2 インターネットタウンページの概要

インターネットタウンページは，NTT の発行する職業別電話帳の情報を基に，掲載情報の検索機能，各掲載と地図や鉄道路線図との連係機能，広告や各種特集記事情報などを WWW から実験提供しているものである [1]．掲載情報の検索は，業種名・企業名・住所の 3 項目の中から任意の個数の項目に対し，検索語として自由語あるいは業種名一覧や住所一覧で提供される統制語を投入するインタフェースになっている．

3 *Best HITS!* の概要

Best HITS! は，インターネットタウンページでのユーザの検索履歴をランキング化したものである．2 章で述べたように，インターネットタウンページでは業種名・企業名・住所という 3 種類の検索条件の項目が存在するが，*Best HITS!* は，次の項目に関するランキングを生成している．

1. 検索された業種名のランキング
2. 検索された業種名と住所の組合せのランキング
3. 2. のランキングを，グルメ，ビジネス，趣味などの業種名のジャンルわけによって，分割したもの (テーマ別ランキングと呼ぶ)

図 1がその画面例である．ランクインした検索語の部分には，実際にその条件で検索を実行するというハイパーリンクを張ってある．このリンクをたどって検索を行うことをここでは，*Best HITS!* Search と呼ぶ．

4 Social Filtering としての検索ランキング生成の要求条件

Social Filtering の一方法として，検索ランキングを生成する場合，ランクインした検索条件を他のユーザに再利用してもらうことになる．[2, 3] などで行われている Social Filtering では，個々の音楽家や Netnews の記事をユーザ自身が数値で評価し，その評価の結果を利用して，ユーザの match making を行って，情報フィルタリングすることが行われている．しかし，多くの一般ユーザに使われる WWW 上の各種検索サービスの場合，個々の検索結果に対し，ユーザにそれら

Best HITS!

☆みんなの検索傾向を参考にすると，とても便利☆
リンクをたどると実際に検索できます！

1997 年 x 月 y 日 現在の 業種＆地域 ランキング

順位	業種 & 地域	前回の順位	Up/Down
1	東京都台東区千束の個室付浴場	1	→
2	東京都台東区のラブホテル	2	→
3	東京都千代田区のビジネスホテル	3	→
4	東京都大田区の家具店	5	▲
5	東京都江戸川区の旅行業	4	▼
6	東京都板橋区の郵便局	6	→
7	東京都千代田区の学校 (高等学校)	10	▲
8	東京都新宿区の中華料理店	11	▲
9	東京都板橋区のレンタルＣＤ・レコード	8	▼
10	東京都北区の司法書士	7	▼

(後略)

その他のランキング
業種ランキング テーマ別ランキング
最新のランキングに戻る バックナンバー

Copyright (C) 1997 NTT

図 1: *Best HITS!* の画面例

の評価を行ってもらうことは，ユーザにかけるコストとしては高すぎる．そこで「何度も検索が行われている検索条件は，他のユーザにも検索が行われる可能性の高い検索条件である」という仮定を立て，検索回数を評価点の代わりに使う方法をここでは採用する．

また，ランクインした検索条件を再利用してもらうためには，少なくとも以下の条件を満たす必要がある．

条件 **1** ランクインした検索条件は，1 件以上の検索結果が得られる検索条件であること

条件 **2** ランクインした検索条件は，実際に，多数のユーザによって繰り返し検索された検索条件であること

これらの条件を満たすためには，単にユーザの検索履歴のみをカウントするだけでは不十分であり，条件 1 に反するような検索条件を排除したり，ユーザの検索履歴のばらつきをある程度，集約するような操作を行ったりする必要がある．特に，インターネットタウンページの場合，自由語による検索が可能であるため，上記のような操作は必須である．

5 *Best HITS!* における検索ランキング生成アルゴリズム

4章で述べたような要求条件を満たすため，*Best HITS!* では，図 2のようにして，検索履歴を加工してランキングを生成している．

まず，ユーザが入力した自由語の種類はあまりに膨大で，各自由語毎の検索回数がほぼ平均して非常に少ない回数になってしまう傾向が実際に見られ，条件 2 に反する可能性が高い．検索システムは，検索時にあいまい検索 [4] の規則により自由語をデータベース上の統制語に展開してから検索を行うため，この展開後の統制語に沿ってランキングを生成する．

また，この規則によると，ユーザの入力した自由語は一般に複数の統制語に展開される．そこで，1 回の自由語による検索をなんらかの基準で個々の統制語による検索として回数を分割してカウントする必要がある．*Best HITS!* では，データベース上の該当統制語

---

* Search Ranking of WWW and its Effectiveness as Social Filtering

† Nobuyuki Miura, NTT Software Laboratories

‡ Katsumi Takahashi, NTT Software Laboratories

§ Ken'ichi Shima, NTT Software Laboratories

[1] http://townpage.isp.ntt.co.jp/besthits/

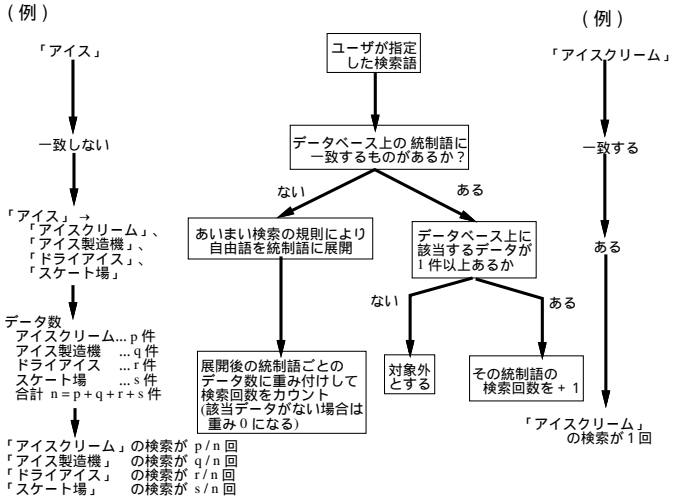

図 2: 検索履歴からのランキング生成

のデータ数を考慮して，データ数の多い統制語に大きな重みをつけてこの分割を行っている．この重みづけは，データ数の多い検索条件ほど，他のユーザによる再利用の割合も高いであろうという仮定に基づいている．また，この重みづけによりデータ数 0 件の検索条件は自動的に対象外となり，条件 1 を満たすことができる．ユーザが自由語ではなく，統制語を入力した場合にも該当するデータ数が 1 件以上の場合のみ，検索回数としてカウントしている．

6 検索ランキングの Social Filtering としての効果の分析

本章では，*Best HITS!* を題材として，検索ランキングの効果を Social Filtering の観点から分析する．

分析対象の概要は，表 1の通りである[2] ．*Best HITS!* を見たのは全ユーザ[3] のうちの約 8% と少なめであるが，これはこのページへのハイパーリンクがトップページから直接張られているものではなく，何段かハイパーリンクをたどった先のページからリンクが張られていて，ユーザの目に留まる割合が低いためと推測される．また，*Best HITS!* を見たユーザの半数が実際に，*Best HITS!* Search を行っていた．

表 1: 分析対象の概要

実験対象期間	提供開始から 49 日間
対象ユーザ数	42124 ユーザ
Best HITS! へのアクセス数 / 全体のアクセス数	日により，2 ~ 3%
Best HITS! からの検索回数 / 全体の検索回数	日により，3 ~ 6%
Best HITS! を見たユーザの割合	8.3% (3429 人)
Best HITS! Search をしたユーザ / *Best HITS!* を見たユーザ	48.3 % (1655 人)

まず，図 3は，ランクインした検索語の順位変動をグラフ化したものである．10 位以下のランキングは時間変動が大きいが，10 位以内，特に 5 位以内のランキングはほぼ安定している．このような順位の安定した検索条件は，他の多くの人に再利用される検索条件となる可能性が高く，Social Filtering としての効果も期待できる．

次に，ランクインした検索条件が情報 Filtering された検索条件としてどの程度利用されているを分析する．

[2] ヒット数のうち，html ファイルおよび html ファイルを生成する CGI へのヒットのみをアクセス数として数えた．

[3] インターネットタウンページでは，HTTP Cookie を使ってユーザの識別をしている [5]

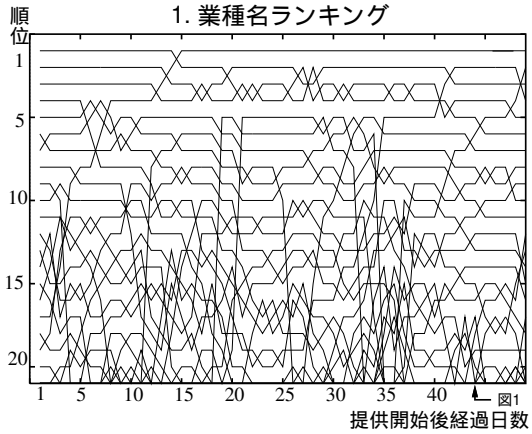

図 3: *Best HITS!* の時間変動

る．実験対象期間中に上位 5 位に一定期間ランクインした検索条件を対象として調べた結果が表 2である．これは，*Best HITS* 導入前に，その業種の検索回数が全体の検索回数の中で占める割合を 1 としたとき，導入後に同じ割合が導入前に比べてどのくらいの倍率で変化しているかを示している．例えば，業種 4 では，導入後はこの業種の検索が全体の検索に占める割合は，1.23 倍と増えているが，通常の検索画面からの検索は 0.90 と 1 割減であり，その増加の理由は，*Best HITS!* からの検索が行われたことによるものである．

業種 1 や 2 では，*Best HITS!* 導入により，*Best HITS!* Search・通常の検索ともに非常に増えており，導入の効果が全体の検索回数増に効果が出ている．業種 3 や 4 では，*Best HITS!* 導入により，通常の検索画面を経ずに，*Best HITS!*の画面上から検索を行う人が増え，Social Filtering としての効果が現れている．業種 5 や 6 には，残念ながら，情報 filtering としての大きな効果は現れていない．

表 2: *Best HITS* 導入による検索回数割合の変化 (倍)

	通常の検索	*Best HITS!* Search	合計
業種 1	5.50	6.60	12.10
業種 2	1.10	2.90	4.00
業種 3	0.83	0.21	1.04
業種 4	0.90	0.33	1.23
業種 5	0.89	0.05	0.94
業種 6	1.10	0.17	1.27

Best HITS 導入前の検索回数の割合を 1 とした

以上から，効果の程度にばらつきがあるものの，Social Filtering の最も簡易な形式である，検索ランキングにもその効果が見られることがわかった．

7 今後の課題

我々は，現在，各ユーザの検索ランキングに相当する個々のユーザのユーザモデルを作成している [6] ．この結果を使って [2, 3] のようなユーザの Match Making による Social Filtering を行う．

参考文献

[1] 島健一, 高橋克巳, 三浦信幸. "インターネット・タウンページの構築 (1) ~概要~". 情報処理学会 第 54 回 全国大会, Vol. 3, pp. 481–482, May 1997. URL: http://townpage.isp.ntt.co.jp/.

[2] Upendra Shardanand and Pattie Maes. "Social Information Filtering: Algorithms for Automating "Word of Mouth"". In *CHI '95*, pp. 210–217. ACM, 1995.

[3] P. Resnick, et al. "GroupLens: An Open Architecture for Collaborative Filtering of Netnews". In *CSCW '94*, pp. 175–186. ACM, 1994.

[4] 高橋, 三浦, 島. "インターネット・タウンページの構築 (2) ~あいまい検索技術~". 情処第 54 回全国大会, pp. 3-483 ~ 484, May 1997.

[5] 三浦, 高橋, 島. "インターネット・タウンページの構築 (3) ~個人適応型 WWW の試み~". 情処第 54 回全国大会, pp. 3-485 ~ 486.

[6] 三浦, 高橋, 島. "個人適応型 WWW のためのユーザモデル構築". In *Interaction '97*, pp. 87–94. 情報処理学会, Feb. 1997.