

## 検索者とサーバーの協調による WWW 情報検索サーバーの性能改善に関する考察

中渡瀬 秀一

NTT情報通信研究所

〒 238-03 神奈川県横須賀市武 1-2356

nakawtse@dq.isl.ntt.jp

### あらまし

近年WWW上には多種多様な情報が存在している。そのためユーザ（検索者）にとって必要な情報はそこから入手できる場合が多い。しかしブラウジングによる検索には多大な手間を要する。本稿では経済的効率を考慮したWWW検索のための情報発信者と検索者との協調について論じ、協調事例を提案する。

キーワード WWW, 情報検索, 経済効率, 協調

## A Cooperative Data Exchange among Clients and Servers for a efficient web navigation

Hidekazu NAKAWATASE

NTT Information and Communication Systems Laboratories

1-2356 Take, Yokosuka, Kanagawa 238-03, Japan

nakawtse@dq.isl.ntt.jp

### Abstract

Many useful resources are produced on the web in recent years, but netsurfer must navigate hard for the information retrieval from the web. This paper describes how a economical efficiency of contents exchange among creators and netsurfers is improved without a competition for resources. Then this paper proposes the cooperative exchange of them for a efficient information retrieval.

key words WWW, information retrieval, economical efficiency, cooperation

## 1 はじめに

現在WWW上には多くの有用な情報が存在しており、それ全体が巨大な辞典、情報の発生源、交換場所のように機能するようになってきている。そのうえサーバー（Webサイト）やコンテンツ作成者も日に日に増加しており、入手可能な情報は増す一方である。ところが現実には我々は限られた時間しかなく、すでに発信されている（興味のある）情報すべてに目を通すこともできなくなっている。そこで情報検索サービスの課題はいかにしてユーザ（検索者）により高い効用<sup>1</sup>をもたらす情報を低コスト（時間、費用）で提供するかということになってきている。

現状でのWWWからの情報の入手はまずURLを知ることから始まる。まだサーバーの少なかったころはほとんどのサーバーや新着ページを一通りアクセスするといったことも可能であったが、すでにそのような方法は不可能になった。そこで今日では次のようなサービスによって検索者はURLを獲得する。

- サーチエンジン 全文検索、キーワード検索によるもの。
- ディレクトリサービス URLをシーケンス分類したもの。
- 紹介 雑誌の紹介記事、ネットニュースやホームページ上の紹介、知人からの紹介など。

### 1.1 問題点

しかし発信されるコンテンツの量が増えるとこのようなサービスにも問題がでてくるようになった。

**サーチエンジン** コンテンツの存在量が膨大なので単純なキーワードではあまりに多くのコンテンツがヒットしてしまう。これを絞り込むためにはキーワードの正規表現や論理結合を使って、検索式をうまく表現しなければならないが、この作業は高度な熟練を必要とする。さらに文字のパターンマッチングによって検索されるため、表記の異なる関連語で表現されたコンテンツが洩れてしまう。

**ディレクトリサービス** 情報提供者側が分類を提示してくれる所以、その階層化された分類に誘導されて欲しい情報までたどり着くのは容易であると思われる。ところが情報提供者側が意図した分類と検索者の意図した分類とが異なった場合にはコンテンツが見つからない恐れがある。またコンテンツを分類するのは情報提供者側の役目であるが、コンテンツの内容を判断して正しく分類することを自動化するのは難しい。さらに新分野の発生などにより既存の分類体系をうまく再編するのはさらに困難である。

<sup>1</sup>ユーザが欲求を充足することによって得る満足

**紹介** 雑誌は全体の概観するのが容易であり携帯に便利という利点があるが、雑誌の種類がすでに膨大になっておりどれを選べばよいか判断するのが難しくなっている。

ここで問題をその性質から2種類に分類してみた。ひとつは限定されたある特定の目標の最適化に関する問題である。この場合、目的を達成するための資源は十分であるとする。そしてもうひとつは希少な資源、手段を用いて、達成したい複数の目的のそれぞれをどのようなバランスで達成するかという問題である。WWW上の情報検索の場合、単一の目標として

1. 検索の対象となるサーバーの数を増やす。
2. 検索速度を向上させる。
3. 適合率、再現率を向上させる。

などがある。この時、目標を限定して例えば検索速度の向上を追求するならこれは前者の問題である。しかし目標が複数あり、これらの目的を達成させるのに必要な資源が競合し合うなら後者の問題も考慮しなければならない。例えば探索範囲の広い検索サービスを提供するために、すべてのサーバーのURLをチェックするには膨大な時間とトラフィックを要する。したがってそのチェックに要するトラフィックとサービス提供に要するトラフィックとが競合することになる。またそのようなサーバーが沢山発生するとサーバー側のチェックに要するトラフィックと検索者のブラウジングのトラフィックとが互いに競合してしまう。ゆえにこの場合、後者の問題が発生していることになる。上の例の場合、解決策は複数のサーバーが同じチェックをするのを避け、互いに必要なものを共有できるようにすることである。本稿では後者のタイプの問題を扱う。

### 1.2 関連研究

関連研究としては例えば網羅的なカタログデータベースを作成する方法[10]、多量のコンテンツを走査して、そこからシーケンスを作成する方法[6, 9]、ページのアクセス確率からアクセスコストを最小化するようなリンク構造を作成する方法[11]、コンテンツ作成者と検索者がそれぞれプロファイルを作成し、それらの照合を常時ネットワーク上で行う方式[4]、個人を表すオブジェクトをネットワーク上に作成し、それらの照合により協調し合う組織を見いだす方式[7]、散在する不特定多数のデータベースに対する問合せ方式[5]などがある。このうち[4, 6, 9, 10, 11]は前者のタイプに属する。[5, 7]では協調について扱っている点で本稿と似ているが、それらが扱うのは主に協調関係が存在するならば機能する協調方式についてであった。本稿では後者の問題としての協調の存在する条件を扱う。そのために経済モデルを用いて情報発信者と検索者との協調について分析し、さらに可能な具体的な協調の例をいくつか提案する。

### 1.3 本稿の構成

以下2では本稿で扱う協調について経済モデルを用いて説明する。そこでは協調を互いの効用を増すような財の交換としてとらえ、その仕組みについて述べる。3ではまず2で用いたモデルから協調の存在する条件を導出し、次に協調の事例を提案し、導いた条件に照らしてそれを検討する。

## 2 協調の効率分析

ここでは効率分析に用いる経済モデル(pure exchange model) [12]を紹介し、それを用いていかなる協調が効率改善をもたらすかについて述べる。

### 2.1 経済モデル

#### 2.1.1 基本的な概念

- 効用関数 UF/utility function)

消費者(検索者)はn種の資源、財、サービス(これらをまとめて財といいう)入手することによって効用を得る。この対応を効用関数として以下のように表す。

UF : 財<sup>n</sup> → 効用

- 無差別曲線 IC(indifference curve)

UFの曲面をある水準で水平に切ったときの断面を無差別曲線といいう。この曲線上の異なる2点では財の配分は異なるがそれらは同じ水準の効用をもたらす。図1に2財の場合のICを示す。

- 限界代替率 MRS(marginal rate of substitution)

ICの微分係数を限界代替率といいう(図1のP\*ではMRS = β/α)。

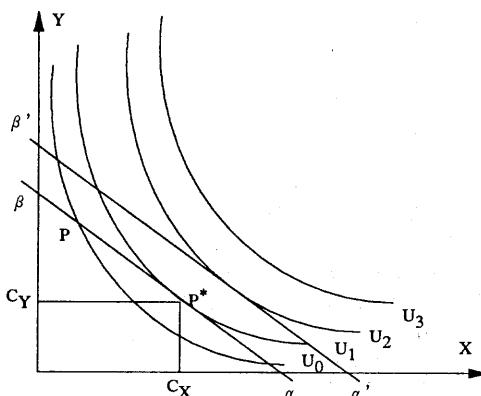


図1: 無差別曲線の例

### 2.1.2 仮定

- 限界代替率遞減性

MRSについては次の仮定をおく、

限界代替率は低下する。 (1)

(1)を限界代替率遞減の法則といいう。これは複数の財の消費を考えるときに特定の財に偏らない消費の方が効用が高いことを表している。視覚的には図1のU<sub>0</sub> < U<sub>1</sub> < U<sub>2</sub> < U<sub>3</sub>様に異なる水準のICが右下がりの曲線群をなすことを表している。

### 2.2 効率的な協調の原理

次に複数の財を配分する際の効率とそれがいかに改善されるかについて説明する。

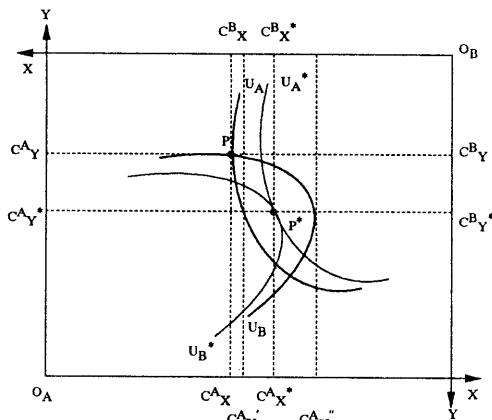


図2: 財の配分効率

図2では二人の消費者A, BのUFのグラフを重ねて描いている。Aのグラフは原点をO<sub>A</sub>から右方と上方にXとYをとり、Bのグラフは原点をO<sub>B</sub>から左方と下方にXとYをとっている。この状態で財X, Yの総供給量をそれぞれC<sub>Xmax</sub>, C<sub>Ymax</sub>とする。この時のP点ではXはAにC<sub>X</sub><sup>A BにC<sub>X</sub><sup>B</sup>、同様にYはAにC<sub>Y</sub><sup>A</sup> BにC<sub>Y</sub><sup>B</sup>配分されており、すべての財がA, Bによって消費されている(C<sub>X</sub><sup>A</sup> + C<sub>X</sub><sup>B</sup> = C<sub>Xmax</sub>, C<sub>Y</sub><sup>A</sup> + C<sub>Y</sub><sup>B</sup> = C<sub>Ymax</sub>)。そしてこの配分のもとでは、A, Bの効用はそれぞれU<sub>A</sub>, U<sub>B</sub>である。ここで別の配分を表す点P\*をみれば、この場合のA, Bの効用はU<sub>A</sub><sup>\*</sup>(≥ U<sub>A</sub>), U<sub>B</sub><sup>\*</sup>(≥ U<sub>B</sub>)となりA, B双方にとってPでの配分より効用が改善されている。このような点の集合を考えると、それは効用がU<sub>A</sub>, U<sub>B</sub>のICによって囲まれたレンズ状の部分になる。特にA, B双方のMRSが等しい点P\*ではこれ以上双方の効用を同時に高めることができないという意味で最適(バレー</sup>

ト効率的)である<sup>2</sup>。つまり複数の消費者が協調してこのような点を見いだし、お互いの財を交換して配分を変更させることによって、財の新規供給をすることなく効率(双方の効用)が改善されるのである。

### 3 情報発信者と検索者との協調

前章では複数の消費者が協調することによって、互いの効用を増加させる仕組みを説明した。本章ではまず情報発信者と検索者がこの仕組みによって情報検索効率を改善できる条件を導き、そのような具体的な協調を提案する。

#### 3.1 効率が改善できるための条件

##### 3.1.1 条件の導出

2.2節で説明した原理によって効率を改善するには以下の問題を解かねばならない。

$$\text{図2における財 } X, Y \text{ として何を選ぶか} \quad (2)$$

$$\text{レンズ状の部分をいかに見いだすか} \quad (3)$$

そこで再び図2を参考にして財X, Y, MRSについて考察する。P点からP\*点へ移行する過程で、まずAがYを $C_Y^A - C_Y^{A''} (= \Delta Y^A)$ だけ手放すとすれば、Aの効用を維持するのには $C_X^{A'} - C_X^A (= \Delta X^A)$ だけXが必要である。しかしBは $\Delta Y^A$ 受け取ることによって、 $C_X^{A''} - C_X^A (= \Delta X^B)$ だけ手放しても効用を維持できる。ここで

$$\Delta X^B > \Delta X^A \quad (4)$$

なので $\Delta X^B > x > X^A$ なるxをBがAに与えることは双方の効用を増すことになっている。 $\Delta X^B, \Delta X^A$ とはB, AそれぞれのXで測った $\Delta Y^A$ の価値であり、その差が正ならばP\*はレンズ状の領域に含まれることになる。これによって(3)の問題を解く方法は以下のようにまとめられる。

- ある財Xで測ったA, Bにとっての財Yの価値(それぞれ $Y_X^A, Y_X^B$ )を計算する

2.

$$Y_X^B > Y_X^A \quad (5)$$

ならばP\*が存在する。

この時AのYとBの $Y_X^B > x > Y_X^A$ なるxとを互いに交換することによってより配分効率は改善される。

<sup>2</sup>本稿では2財2人のモデルで説明しているが、k人n財の場合にも最適点が存在することが証明されている[13, 14]。

##### 3.1.2 条件を満たす事例

次に得られた条件を満たす事例について説明する。ここではXとして特に貨幣をYとして何か商品を考える。またn単位のX, Yをn<sub>x</sub>, n<sub>y</sub>と表記する。消費者A, BにとってXで測った商品の価値がそれぞれ

$$10x = 50y, 10x = 100y$$

の関係として表される。すると $100y > 50y$ であるから例えばAが10x払って70yだけBから受け取るとA, B双方にとってそれぞれ20y, 30yに起因するだけ効用が増加したことになる。ちなみに今の場合Xは貨幣であるから、Yの1単位に対するXの量は価格ということになる。

### 3.2 協調事例の検討

本節ではA=検索者, B=情報発信者として双方にとって効率的な協調の事例をいくつか提案し、それについて検討する。

##### 3.2.1 効率の検討

まず交換の対象となる財を検討する。そのような財は2種類あり、AがBに与える財とBがAに与える財である。それぞれをX<sub>AB</sub>, X<sub>BA</sub>と表す。現在X<sub>AB</sub>としては使用料、(検索サービスに掲載されている)広告の検索者による視聴、X<sub>BA</sub>としては検索(情報提供)サービス、がある。検索サービスと使用料の組み合わせで交換が成立しているサービスは有料データベース、有料の電子メールニュース[1]などであり、検索サービスと広告の視聴の組み合わせでは、無料のサーチエンジン、ディレクトリサービス[2]がある。上記以外のX<sub>BA</sub>として、ここでは検索者の検索記録や検索評価を検討する。

##### 3.1.1節の結果によれば検討すべきものは

- 検索者と情報発信者からみた検索サービスに対する検索記録や検索評価の価値

である。この差(5)が存在すれば、理論的には財の交換によって効率が改善されることになる。

検索者からみた場合 検索記録は通常クライアントやサーバーが自動的に行う<sup>3</sup>ので検索者がそのためには追加操作をする必要はない。そのため検索者側にかかる負担は少ないと思われる。また検索評価とは検索者の問合せと検索結果に対する適合集合<sup>4</sup>や再現集合<sup>5</sup>の組みがそうであるが、適合集合の方をチェックするには容易なことだと考えられる。つまり

<sup>3</sup>proxy 経由接続するなどして検索者のプライバシは配慮されなければならない

<sup>4</sup>サーバーの検索結果のうち必要な(役に立った)もの

<sup>5</sup>サーバーの検索結果として期待していたもの

**S 1** : 検索者からみたときの検索サービスの価値（検索サービスに出力として期待するものを自ら作り出す手間）のほうが検索記録や検索評価の価値（それを作る手間）より大きい。

**情報発信者からみた場合** 検索記録によって得られる統計から検索頻度の高い問合せ（サーチエンジンの場合）または分類項目（ディレクトリサービスの場合）がわかる。これは視聴率調査に相当するもので広告スポンサーの募集に有用である。さらに得られた統計の性質によっては以下のような検索機能の改善が期待できる。

**ディレクトリサービスの場合の検索速度** 検索頻度の高い分類項目に対するサーチパスを短くするよう分類構造を再編成し、検索速度を向上させる[11]。

**問合せ全体に対する平均的な再現率** 検索頻度の高い分野のURLの収集量を強化する（より多くのサーバーにアクセスしてURLを収集する。URLの収集を頻繁にする）ことによってこれによって平均的な再現率が向上する。

**問合せ全体に対する平均的な適合率** 適合集合にヒットしているURLに対し検索結果ランキングを高くすることによって平均的な適合率が向上する。

このように情報発信者にとっての検索記録のような統計情報は有用である。また

**S 2** : 情報発信者にとって自ら生産できないという意味において、統計情報は検索サービスよりも価値があると考えられる。

S1,S2より情報発信者にとっての検索記録や検索評価の価値は検索者のそれよりも大であると思われる。したがって(5)が成立し、検索者と情報発信者との間に互いの効用を改善する財の交換（検索サービスと検索記録や検索評価）が存在する。

### 3.2.2 協調の具体例

前節では財の協調的な交換によって効率を改善できる可能性について示した。本節では事例のより具体化した例を説明する。

**検索記録の取り方** 検索記録の場合には情報発信者側が検索記録を取得するシステム<sup>6</sup>を用意する。

**検索評価と交換する検索サービス（交換用検索サービス）** 検索評価と交換する検索サービスは無料で提供されている検索サービスよりも便利である必要がある。このようなサービスを考えるために検索者の利便を分類すると1:検索機能、2:時間、3:費用に関するものが考えられ

<sup>6</sup>既存のサーバーの多くはアクセスログの記録を行う機能を持っていている

る。これらの利便を提供できるようなサービスを以下に示す。またそれぞれの分類の中でも異なる特徴を持つサービスが考えられる。それらによって検索者のより多様な嗜好に対応することができる。

### 検索機能の優れたサービス

#### 1. 多くのサーバーの検索

2. 複雑な検索処理 キーワードの（部分）マッチングのみならず、統計的シーケンス[9]などによる関連語を含めた検索。

3. 最新情報 コンテンツは情報発信者によって自由に作成、更新、削除されるため最新の情報を常に保持するためにURLの収集を頻繁に行う。

**時間を節約するサービス** 技術的にはハードウェアの性能の向上によっても検索時間も短縮されうるが、ここではサーバー側の性能に依存しない方式を示す。

● 優先的処理 現在、一部のサーチエンジン[3]では過負荷のために処理を受け付けられない状態になることがあるが、このような場合に優先的に処理させるという権利を与えることが考えられる。これによって検索のための待ち時間が短縮されるため、これはみかけ上、検索速度が向上したのと同等の効果を産む。

**費用を軽減するサービス** 高機能な検索を要求せず、時間にも余裕のある検索者にとっては費用の負担の小さいサービスは魅力的であると思われる。内容としては検索者の負担する費用を軽減することがあげられる。

これらのサービスは例えば次のように交換される。

#### [検索サービスと検索記録の交換手順]

1: ユーザがサーバーにアクセスする。

2: サーバーが検索サービスと検索記録や検索評価の交換を求める。

3: ユーザが承諾した場合には交換用検索サービスが提供され、サーバーは検索記録を取得する。

## 4 終わりに

本稿では情報発信者と検索者の協調を分析する方法について説明し、それによって効率を考慮した情報発信者と検索者の協調の存在について検討した。本検討の特徴は経済的効率を対象にしていることになり、検索者と情報発信者とを合わせた全体としての効用の改善が分析できることである。また可能な協調の具体例について触れ、それぞれの特徴を示した。今後はその他の協調の検討も同様に行いたい。

## 参考文献

- [1] <http://www.watch.impress.co.jp/>

- [2] <http://www..yahoo.com/>
- [3] <http://www.info.wasda.ac.jp/search.html>
- [4] 松本宏, 植村俊亮, “インターネットによる能動型情報提供システムの提案”, 信学技報 DE96-8, pp.43-48, 1996.
- [5] 宮崎収兄, “不完全データベースと広域データベース検索”, 情処研報 DBS106-18, pp.131-138, 1996.
- [6] 佐々木寛, 羽生田博美, 木下哲男, “単語表記情報に基づく情報検索向き単語関係の抽出”, 信学技報 AI95-7, pp.9-16, 1995.
- [7] 木實新一, 上林弥彦, “ネットワーク上の仮想組織における Deputy オブジェクトの利用”, 信学技報 DE95-41, pp.25-31, 1995.
- [8] 横田一正, 木實新一, “ネットワーク上の仮想データベースシステム”, 信学技報 DE95-43, pp.41-47, 1995.
- [9] 城風敏彦, 羽生田博美, 木下哲男, “統計的シソーラスを用いた分散型ネットワーキュース検索システム”, 信学技報 AI95-24, pp.15-22, 1995.
- [10] 林良彦, 菊井玄一郎, 鷺崎誠司, 砂場倫太郎, “WWW 情報空間における Resource Discovery と Navigation 支援”, 信学技報 AI95-31, pp.71-77, 1995.
- [11] 高田喜朗, 辻野嘉宏, 都倉信樹, “アクセスコストを最小化するハイパーテキストの構成法”, 信学技報 SS95-22, pp.17-24, 1995.
- [12] 西村和雄, “ミクロ経済学”, 東洋経済新報社, 1995.
- [13] K. J. Arrow, G. Debreu, “Existence of Equilibrium for a Competitive Economy”, *Econometrica*, Vol.22, pp.265-290, 1954.
- [14] L. Mckenzie, “On Equilibrium in Graham's Model of World Trade and Other Competitive Systems”, *Econometrica*, Vol.22, pp.147-161, 1954.