

## WWWの情報取捨選択行動におけるインデックスの影響分析

溝渕 佐知(NTT<sup>1</sup>)

古山恒夫(東海大学<sup>2</sup>)

buchiko@slab.ntt.co.jp furuyama@wing.ncc.u-tokai.ac.jp

### 要 約

ネットワーク上で流通する情報は増加の一途を辿っているが、人間の情報処理能力は無限ではなく、それらの情報を有効に活用するための支援が必要とされる。本研究は、WWWにおいて、人が多くの検索結果の中から、より適切なものを効率的に選択するための支援技術に関するものである。

本稿では、情報選択手がかりとしてのインデックスの利用に着目し、その有効性、及びインデックスの提示方法による選択パフォーマンスへの効果を調べる実験について報告する。実験の結果、イメージベース及びテキストベースともその有効性が確認されたが、両者に有意な差は見られなかった。また、一部の課題で、イメージの同時提示数による効果が見られた。

## An analysis on the effect of indexes on WWW information selection process

Sachi Mizobuchi (NTT) Tsuneo Furuyama (Tokai University)

### Abstract

The amount of information on the network is constantly increasing. However, human capacity of information processing is limited, so that people expect to have some support tools to handle such growing size of information over the Internet.

This study aims to find the technology that helps people to get suitable information efficiently when they face a big number of search results on a web site search engine.

In this paper, we focus on the use of indexes as a cue to select appropriate information. We conducted two experiments to understand about the validity of index usage and the effect of presentation style on subjects' performance. The result of experiments shows that both text-based and image-based indexes work without any significant differences. We also found that, in some sorts of tasks, the number of indexes in a scene controls on selection performances.

<sup>1</sup> NTT ソフトウェア研究所 180-8585 武蔵野市緑町 3-9-11

<sup>2</sup> 東海大学開発工学部 410-0395 静岡県沼津市西野 317

## 1. 背景

### 1.1. WWW における情報量の増大

現在、WWW 上に存在するページは全世界で 3 億以上にものぼると推定されており[1]、インターネット人口の増加に伴い、今後も発信され蓄積される情報はますます増加し続けることが予測される。一方で、人間の情報処理能力は無限ではなく、扱う情報量が増せば、その処理に伴う負荷も増大し、その結果、取捨選択のパフォーマンスは低下すると考えられる。「情報」は人によって解釈されて初めて「知識」として利用可能になり、価値創造に結びつく。したがって、ネットワーク上の情報が増大してもなお有効に活用されるために、人が大量の情報の中から求めるものを見つけ出すという課題に伴う負荷を軽くするための支援が不可欠である。このためには、人が情報を取捨選択する際の認知プロセスについて理解し、それに合った支援法を開発・適用することが望まれる。

### 1.2. 情報の取捨選択の負荷を低減させるための二つのアプローチ

情報の取捨選択に伴う負荷を低減させるためには、システム側からのアプローチと人間側からのアプローチの二つが考えられる。すなわち、

- (1) 予めできる限り妥当な情報を、自動的に絞り込むことに重点を置く (フィルタリング・アプローチ)

これは、結果として提示する情報を選択する機構をよりインテリジェントにし、絞り込みの精度を増すことを指向するものである。情報の削ぎ落としを可能な限り自動化し、実際に人間が扱う選択肢を極力小さくすることがゴールとなる。この中には、他者や自分が過去に行った検索の結果やプロセスを利用するというアプローチが含まれる[2][3][4]。

- (2) 人間が絞り込みをしやすいように支援することに重点を置く (選択支援アプローチ)

手がかりとして利用可能な情報を付加する、あるいは余分な情報を削減する、などの手段によって取捨選択の負荷を下げる。意思決定支援ともいえる。

よりよい情報環境の構築のためには、これらの二つのアプローチを両方進めていく必要があると考えられるが、本研究は、このうち (2) のアプローチを指向するものである。

### 1.3. WWW での情報検索のタイプ

WWW には、公共機関のサービス情報、企業の製品紹介や採用情報、電車やバスの時刻表、大学の講

義内容、個人の日記やエッセイなど、多様な情報が発信されており、情報を検索する側の意図も様々である。雑多な検索目的を、収集した情報の利用という観点から大別すると、特定の情報を得た時点でニーズが満たされる「確認型」と、次のアクションへのきっかけとして、つまり手段として情報を収集する「探索型」とに分けられる。確認型では、表現は様々であっても、要求される内容は一つである。例えば、「あるイベントが何月何日に行われるか」が知りたい場合、複数のページにそのイベントについての情報が様々な形式で掲載されている可能性はあるが、そのうちで、開催日時の記載されたものを一つ見つければ後のものは見る必要が無い。この場合、吟味されるのは開催日時という情報の有無のみであり、それが記載されているウェブページ群の中では各ページの有効性は検索者にとって等価であると考えられる。

一方、探索型は、答えは一つに決まらない、あるいは、答えがあるかどうかさえ分らない場合があることを前提として行われる検索である。例えば、「コミュニティ」というキーワードでの最近の研究動向が知りたい場合、表現、内容ともに「よりよいもの」が存在する可能性を持っている。このような性質を考えると、探索型においては、検索者が選択肢の中からより妥当なものを判断して選び出す、という過程が必須といえる。

### 1.4. 情報選択の手がかりとしてのインデックス

より妥当なものをより速く、かつ正確に選び出す手段の一つとして、元の情報の内容を代表するような抽象化された手がかり (インデックス) を利用することが考えられるが、WWW においてどのような表現形式のインデックスが有効であり、そこではどのような情報内容が代表されている必要があるのかについては現段階で明確になっていない。また、先行研究から、問題解決場面における選択肢の数によって、最適な問題解決方略が変化することが知られており[4]、有効なインデックスを考える上では、同時に提示される選択肢の数についても検討が必要である。

## 2. 目的

本研究は、探索型の検索において、解の候補として挙げられた選択肢の中から、より妥当なものを、速くかつ正確に選び出すための方法について明らかにしようとするものである。今回は実験を通じて、WWW におけるインデックスの有効性の検証、及びインデックスの提示方法による選択パフォーマンスへの効果を調べる。

効果検証のために使用するインデックスのタイプ

として、本実験では、タイトルを示す文字列とウェブページ内の文字列が先頭から一部表示されるもの（テキストインデックス）と、ウェブページの画面そのものを縮小させたもの（イメージインデックス）とを使用する。イメージインデックスについては、サイズ（縮小率の違いによって変化させる）と同時提示数を変化させ、それぞれの要因の効果を調べる。

なお、選択パフォーマンスは、選択肢の中から最終的にいくつかの情報に絞り込むまでに要した時間（以下「選択時間」と、選択されたものの妥当性と）の2つの側面から検討する。

2つの実験を実施する。まず実験1では、イメージ形式のインデックスについて、同時提示数およびサイズの効果を検討する。次に実験2では、主に表示形式（イメージ/テキスト）の効果を検討する。

### 3. 実験

#### 3.1. 実験1

##### 3.1.1. 目的

イメージ形式における、インデックスの表示サイズと同時提示数の効果を調べる。

##### 3.1.2. 方法

表1の○のついたセルにあたる条件でWWW上の情報のインデックスを被験者に提示し、選択パフォーマンスを測定した。

表1 条件の割りつけ

同時提示数	サイズ		
	大	中	小
4	○		
8		○	
16			○

上記の要因配置では、サイズと同時提示数が同時に変化しており、それぞれの主効果が見られないという問題があるが、実験1では、探索的なアプローチを取ることにした。つまり、実験1で最も良好なパフォーマンスの見られる条件を抽出し、実験2において、この条件においてサイズと同時提示数の効果を別々に検討できるように要因を配置する。

インデックスから情報を選択する課題を3種類用意し、一人の被験者が3つの条件（4-大、8-中、16-小）を、それぞれ異なった課題で行なうようにした。課題と条件の組み合わせ、課題・条件の順序については、後述の被験者グループ内で出現頻度が同じになるようにランダムに配置した。

【被験者】東海大学の学生12名。実験への協力依頼時に、質問紙により、各人のWWW利用経験、

WWWでの検索サービス利用頻度、検索サービス利用スタイルを調査し得点化した。これに基づき、WWW利用経験が全く/ほとんど無い者は除いた上で、得点がグループ間で均質になるよう3つのグループに振り分けた。練習試行中に、全員がマウス操作およびブラウザの扱い方については習熟していることが確認された。

#### 【課題】

課題1 肩凝り：肩凝りに悩んでいる友人のために、肩凝りに効く、家でできる運動について知るのに役立つと思われる情報を探す。

課題2 携帯電話：携帯電話を新しく購入しようとしている友人のために、携帯電話を安く買うために役立つと思われる情報を探す。

課題3 京都観光：週末に京都を半日で観光したいと思っている友人のために、観光プランを立てるのに役立つと思われる情報を探す。

【装置・環境】富士通社製パーソナルコンピュータ（FMV DeskPower SV7267, SV237）、17インチディスプレイ、録画用カメラ各2台。東海大学開発工学部キャンパス内の2つの個室にそれぞれ一台ずつ、コンピュータ、ディスプレイ、カメラを設置した。被験者、実験者の位置も含めた配置は図1の通り。刺激提示にはブラウザ（Microsoft社Internet Explorer 4.0）を用いた。

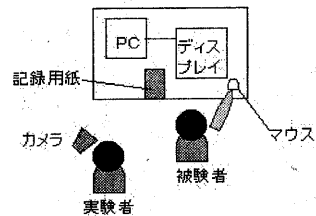


図1 配置図

#### 【材料】

イメージインデックス 既存のWWW上の情報検索サービス[5]で、以下のそれぞれの単語をキーワードとして入力し、検索結果として得られたURLについて、アクセスできないもの、重複したものを除いた上位32件を使用した。

- ・肩凝り：「肩凝り」「運動」 ヒット件数=420
- ・携帯電話：「携帯電話」 ヒット件数=52642
- ・京都観光：「京都」「観光」 ヒット件数=21080

700×700ピクセルの大きさ（元サイズ）でブラウザに表示したものを静止画像としてキャプチャーしたのち、148×148（小サイズ）、223×223（中サイズ）、348×328（大サイズ）ピクセルの3つのサイズに縮小した。これらは、17インチディス

レイ画面に、それぞれ 16 枚、8 枚、4 枚表示できる最大の大きさになるよう調整したものである。縮小イメージから元サイズの画像にリンクが張られ、縮小イメージをクリックすることによって元サイズの画像が見られるようになっていた。

4 枚条件は、大サイズを縦 2 列×横 2 行ずつ並べて 8 画面、8 枚条件は中サイズを縦 2×横 4 ずつ並べて 4 画面、16 枚条件は小サイズを縦 4×横 4 枚ずつ並べて 2 画面用意し、前後の画面には、各画面に用意されたリンクをたどってあるいはブラウザの機能を用いて移動できるようにした。

なお、本稿では以下、元サイズの画像を情報本体とし、インデックスからリンクを辿って元サイズの情報をブラウザに表示させることを「情報の参照」と呼ぶ。

【手続き】図 2 は実験の流れを示したものである。課題の設定のために行った教示は以下の通りであった。

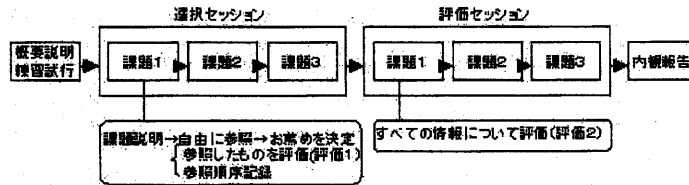


図 2 実験の流れ

課題 1: 「肩凝りに悩んでいる友達があり、家でできる、肩凝りに効く運動はないかと相談されました。WWW で役立ちそうな情報を探して、教えてあげたいと思います。ここに、ある検索エンジンで検索した結果の一覧があります。これら(インデックス)をクリックすると大きな画面が見られます。インデックスの一覧の中からよさそうなものを自由に開いて、一つインデックスを開く(=情報を参照する)度に、参照順序と評価点を記入して行ってください。」

課題 2, 3 についても同様に教示を行った。

評価基準は以下の 5 段階であった。

1. 全的の外れである。
2. テーマに関連がないわけではないが、あまり参考になりそうにない。
3. テーマに関連はあり、もしかしたら参考になるかもしれない。
4. かなり参考になると思うが、今一步である。
5. これは非常に参考になる。

評価セッションでは、全ての情報を参照し、評価点をつけるよう教示した。この際、選択セッションで行った評価にとられることなく、できるだけ客観的な目で再度評価するよう求めた。

### 3.1.3. 結果

まず、被験者グループ間で、選択時間に差が無いことを確認した。12 名の被験者のうち、「すべての情報を前から順番に参照する」という選択方略を取った 1 名のデータについては分析対象から除外した後、条件間で選択時間の差を比較したところ、条件による選択時間の有意な差は見られなかった。しかし、11 名中 8 名について、8 枚-中サイズ条件の選択時間が最も短かったこと、また、実験終了後に被験者より「8 枚-中サイズ条件が最も選択しやすかった」とする内観報告が 12 名中 7 名見られた(4 枚-大サイズ条件 3 名、特に言及なし 2 名)ことを考慮し、実験 2 では、8 枚-中サイズ条件とテキスト条件の比較、および 8 枚条件、中サイズ条件の主効果を調べることにした。

詳細は実験 2 の結果を含めて後述する。

## 3.2. 実験 2

### 3.2.1. 目的

- ・サイズと同時提示数の主効果(実験 1 で見られた、8 枚条件の優位性が、主にサイズの効果によるのか、同時提示数の効果によるのか)を調べる
- ・イメージインデックスとテキストインデックスの間のパフォーマンスの差を調べる

### 3.2.2. 方法

【被験者】東海大学の学生(11 名)および職員(1 名)計 12 名。実験 1 と同様、得点がグループ間で均質になるよう被験者を 3 つのグループに分け、各条件がグループ内で同頻度で出現するよう、条件を割り振った。

#### 【材料】

・イメージインデックス 実験 1 で使用した 8 枚-中サイズ条件の他に、新たに 4 枚-中サイズ条件を作成した。これは、中サイズのインデックスを縦 2×横 2 ずつ並べたものを、計 8 画面用意したものであった。

・テキストインデックス 実験 1 の材料の基となった検索結果一覧の、タイトルと内容の一部をテキストによって表示するタイプのインデックスを作成した。1 画面に 8 件×4 画面、計 32 件の情報のイン

デックスを表示した。タイトル部分をクリックすることによって、実験1で使用したのと同じ、元サイズの画像が見られるようになっていた。

### 3.2.3. 結果

以下、実験1の結果と合わせて検討する。

#### (1)「よい情報」はどこにあるか?--評価2の分布

データ：24名の被験者の、評価2（第2セッションでの評価）の値と、検索エンジンの出力した順序との間に有意な相関関係は見られなかった。つまり、少なくとも上位32件の中からより妥当なものを選択するための手がかりとしては、検索エンジンの算出する「信頼度」はそのままでは有効に機能せず、別の手がかりが必要と考えられる。

#### (2)インデックスの有効性について

インデックスの利用によって妥当なものが選択できていたかどうかを調べるため、全参照パターンの被験者のデータを除外した上で、全ての情報に対する評価2の平均値と、参照されたもののみに対する評価2の平均値の差を検定した。参照されたものの評価値は、全体平均よりも有意に高かった。イメージ/テキスト別に見ても、いずれも参照されたものの評価値は平均よりも有意に高かった（表2）。なお、イメージ形式とテキスト形式の、参照されたものについての評価2の平均値には有意な差は見られなかった。

個人別に見た場合、22人×3課題の66試行のうち、28試行について、参照されたものの評価値がプラス方向に有意な差があり、マイナス方向で有意な差のあるものはなかった。これらの結果より、本実験で使用したインデックスは、より評価値の高いもの、つまりより妥当なものを選択するための手がかりとして有効に機能していたといえよう。

表2 評価平均値の比較

	全体の評価平均値		参照されたものの 評価平均値
全データ	2.67(N=2208 分散1.46)	<	3.03(N=883)
イメージ	2.67(N=1822 分散1.49)	<	3.01(N=757)
テキスト	2.68(N=384 分散1.33)	<	3.16(N=126)

#### (3)選択時間に対する各要因の効果検証

全情報を順次参照した被験者のデータ（2名分）、特異値と考えられる被験者のデータ（2名分）を分析対象から除いた<sup>3</sup>上で、選択時間に対する以下

<sup>3</sup> 選択時間が最も長かった被験者と最も短かった被験者について、残りの被験者の選択時間の平均値との比較を行った結果、この二人についてはその他の被験者との間に有意な差が認められた（ $\alpha = 1\%$ ）ため、特異値と見なした。

下の要因の効果を検証した。

#### ・同時提示数の効果

イメージリストの中サイズ条件のデータについて、課題と同時提示数による選択時間の差を分散分析によって検定した。この結果、課題と同時提示数の交互作用が有意であったため、課題別に選択時間の差を検討することにした。

図3は、各同時提示数条件における選択時間の平均値を課題別に表したものである。

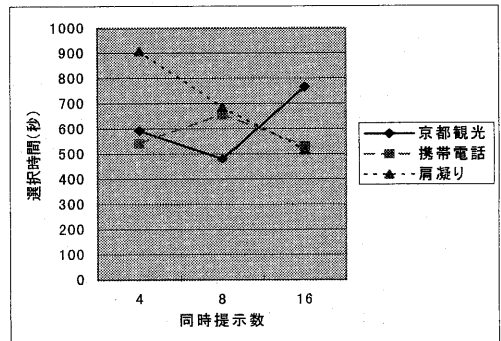


図3 課題ごとの選択時間の平均値

課題3（京都観光）についてのみ、同時提示数の有意な効果が見られた。さらに下位分析を行ったところ、8枚条件と16枚条件との間に有意な差があることが分かった。

#### ・大きさの効果

実験1、2のイメージ条件のデータについて、課題とサイズによる選択時間の差を分散分析によって検定した。この結果、サイズの効果は有意には見られなかった。また、課題とサイズの交互作用も見られなかった。

#### ・表示形式の効果

イメージ条件とテキスト条件との間に、選択時間の有意な差は見られなかった。

#### (4)提示順序と参照順序との関係

全て参照した被験者のデータを除いて、実験1、2のイメージ条件において、提示数毎に、提示順序との相関係数を求めたところ、4枚、8枚、16枚条件でそれぞれ0.58,0.52,0.39であった（いずれも1%水準で有意）。

全体として、参照順は画面の順序に大きく影響を受けるが、同時提示数の増加に伴い、画面内では前から順に見ていく傾向が薄れ、参照順序がランダムになってゆく傾向が認められた。

<sup>4</sup> ピアソンの積率相関係数による

## 4. 考察・展望

### 4.1. インデックスの機能について

3.2.3. (2)で、イメージインデックス、テキストインデックスいずれについても、インデックスを利用することによって、全く無作為に同数の情報を選択するよりも妥当なものを選択できることが示され、インデックスの有効性が確認された。本実験では、2つの表示形式の間に差は見られず、「よりよいインデックス表示形式」を明らかにするには至らなかったが、被験者からは、イメージにテキストによる短い要約がついた形式のものがあれば、より選択がしやすいのではないかとの報告が得られており、今後検討を進める必要がある。どちらか一方の形式のみより複合形式が望まれる背後には、表示形式によって、人に伝わる情報内容が異なることが考えられる。それぞれの形式の利点と、その利点が活かされる状況について明らかにすることによって、両者が最も効果を発揮するような利用法を明らかにすることが出来ると考えられる。

### 4.2. 同時提示数の影響について

本実験では、課題によってパフォーマンスに違いが見られた。これは、各課題で提示された32枚の情報セットに、何らかの違いがあったためと考えられる。違いの詳細については、課題間の情報セットについての質的な検討が必要であろう。

また、同時提示数の変化は、参照順に影響を与えることが明らかになった。この結果から、参照順をコントロールするための要因の一つとして同時提示数があることが示唆される。

### 4.3. サイズと情報量の関係について

今回は「インデックスのサイズ（縮小率）」を変数として統制し、結果、有意な差は見られなかったが、今後は、縮小率が変わることによって被験者にとっては何が変化しているのかを詳細に検討する必要があるだろう。ディスプレイの表示能力には限界があり、画像の縮小率を上げると、小さな文字は潰れて見えなくなり、内容が判別しにくくなる。こうした、見える（＝利用可能な）情報量の変化は、人の情報取捨選択行動に影響を与えると考えられる。今回実験で使用したインデックスは、サイズに応じてほぼ以下のように内容の見え方が異なっていた。

・小サイズ：本文の文字は全く読めず、タイトルが  
かろうじて見える（大きな文字のもののみ）、また  
写真や図表などの存在が分かる分かる程度

・中サイズ：タイトルと本文の見出しが見える程度  
・大サイズ：本文の内容まではほぼすべて見える程度  
ただし、これらは元々のホームページのデザイン（レイアウト、字体、配色など）に依存するため、縮小率がそのまま見える情報の変数を反映しているとは言えない。情報量の影響を検討するには、縮小率ではなく、別の指標によって条件を統制した上で各条件間の比較を行う必要がある。

### 4.4. 参照パターンについて

本実験の被験者において、全情報を順次参照、画面は順序通りだが画面内はランダムな順序で参照、画面間も画面内もランダムな順序で参照、一度最後まで見て、もう一度前に戻る、最後まで見て後ろから前に戻る、などといった、複数の参照パターンが観察された。これらのパターンは、インデックスの総量が増えたり、時間的制約が課されるなど、負荷のかかった状態でも同様に見られるのかどうかは、今後の検討が必要である。

例えば数千件の解候補（検索結果）が提示された場合、それらの内容を全て見た上でよいものを決めるという方略が現実には生じるとは考えにくい。その場合、全参照型はランダム型に移行するのか、一定の範囲内で全参照を行うのか、あるいはそうした状況では全く新しいパターンが出現するのか、様々な条件下で検証する必要があるだろう。また、参照パターンには個々の被験者の認知型が影響している可能性がある。今回は被験者の個人差を考慮の対象外として分析を行ったが、被験者特性と参照パターンの交互作用を調べていくことで、人の情報取捨選択行動がより詳細になると考えられる。

## 5. 参考文献

- [1] Lawrence and Giles (1998) Searching the World Wide Web, Science vol.280 pp98-100
- [2] Collaborative Filtering  
<http://www.sims.berkeley.edu/resources/collab/>
- [3] 協調フィルタリングに関する研究動向  
[http://cairo.aist-nara.ac.jp/~tomohi-f/Docs/cf\\_review/cf\\_review.html](http://cairo.aist-nara.ac.jp/~tomohi-f/Docs/cf_review/cf_review.html)
- [4] 森田昌宏 速水治夫 (1996) 情報フィルタリングシステム 情報処理 vol.37, No.8, pp751-758.
- [5] goo <http://www.goo.ne.jp>
- [6] Payne, J. W., Bettman, J. R., Johnson, E.J. (1993) The Adaptive Decision Maker Cambridge University Press