サイト評価情報を用いたWWW検索表示方法

灰原清太郎† 川越恭二††

利用者が必要な情報をWWWで効率よく得ることができるよう，多くの検索サービスが運営されている．その1つの方法として，予め設定したカテゴリ構造でWWWサイトの名前・URL・紹介文などのデータを人間の手によって分類し，このカテゴリ構造を利用者が用いて検索するディレクトリサービスがある．しかし，増大するWWWに応じて，ディレクトリサービスの登録サイト数も増加している．このため，1つのカテゴリに多くのサイトが含まれ，利用者の操作コストも増えるという問題がある．そこで，サイトごとの評価情報を算出し，その評価情報を用いて，サイトを並び替えて表示する方法を提案する．本方法では，バランスレシオと呼ぶ変数を用いて利用者自身で並び替えの制御を行うことができる．

Site-score based Ordering for WWW Search Result.

SEITARO HAIBARA† and KYOJI KAWAGOE††

There have recently been many WWW search services in order for Internet users to easily find appropriate WWW sites. Some of the search services are providing a sort function on which WWW sites after searching are sorted by some site evaluation scores such as register date or keyword matching degree. It becomes more easily for users to check many of the result sites after sorting. However, the sorting score is only selected among many alternative scores. The paper presents the balance ratio, representing sorting requirement by the user. Using the balance ratio, the total site score for each resulted site is calculated by three of the scores: access score, update score and information contents score. The evaluation result shows that site ranking by sorting the total score using the balance ratio varies with its balance ratio value. With this sorting mechanism using balance ratio, users can easily control WWW search result is obtained.

1. はじめに

利用者が必要な情報をWWWで効率よく得ることができるよう，多くの検索サービスが運営されている．その1つの方法として，予め設定したカテゴリ構造でWWWサイトの名前・URL・紹介文などのデータを人間の手によって分類し，このカテゴリ構造を利用者が用いて検索するディレクトリサービスがある．増大するWWWの数に応じて，ディレクトリサービスの登録サイト数も増加しているため，1つのカテゴリには多くのサイトが含まれるようになる．1つのカテゴリ内では，紹介文などを頼りに興味に合うサイトを検索していくが，調べるサイト数が多くなると，ここでの利用者の操作コストが増える．そこで，サイトごとに評価情報を設定し，その条件で並び替えて表示する方法を提案・評価する．なお，本方法では，バランスレシオと呼ぶ変数を用いて利用者自身で並び替えの制御を行うことができる．

本論文では，第2章では実際のディレクトリサービスを説明し，その問題点を述べる．それをふまえて第3章では，提案した検索表示方法の目的について述べる．第4章では，提案した検索表示方式の概要について述べる．第5章では，提案した検索表示方法の評価を述べる．最後に，第6章では，この研究の今後の課題と期待を述べる．

2. 現在のディレクトリサービス

ディレクトリサービスのためのWWW検索表示方法を検討するために，現在，運営している検索サービスについて，その問題点などを把握する必要がある．

2.1 検索サービス・ディレクトリサービスの方式について

インターネットで提供されている検索サービスは，ディレクトリサービスと検索エンジンの2種類に大きく分類できる．

検索エンジンは，収集ロボットやスパイダー[12]と呼ばれるWWW探索プログラムが，WWWサーバ上のページ情報を定期的に自動収集する．自動的にデータ

† 立命館大学大学院 理工学研究科
†† 立命館大学 理工学部
Ritsumeikan University.
1-1-1, Noji-Higashi, Kusatsu, Siga, 525-8577, Japan.

を収集するため情報量が多いが，情報検索に不慣れな初心者などにはキーワードによる検索が分かりにくく，キーボードによる文字入力の操作コストも高いという欠点がある．

一方，ディレクトリサービスは，WWWサイトについてのサイト名・URL・紹介文などのデータを予め設定したカテゴリ構造に人手によって分類する．カテゴリ構造内のカテゴリ名を頼りに，自分の興味に近いカテゴリを順にたどり，カテゴリ階層を掘り下げていく方法である．この方法は，初心者でもなじみやすい．このため，ディレクトリサービスは，WWW利用の定番的な行動の1つにもなっている．

2.2 ディレクトリサービスの問題点
– サービス提供者の負担

ディレクトリサービスは，検索エンジンと比較すると，ディレクトリサービスを運営・提供する側（以下，サービス提供者）の人手に頼ることが多い．ディレクトリサービスでは，サイト運営者からの依頼を受けて，データベースに登録する．利用者に効率的なディレクトリサービスを提供するためには，サービス提供者が依頼を受けたサイトを適切なカテゴリに振り分ける必要がある．振り分ける人間が，依頼を受けたサイトのジャンルなどを正確に判断できるかどうかも，重要である．ほかに，登録時には，サイト運営者からの紹介文を適切な文に書き換えたり，URL変更・デッドリンクのチェック管理など，サービス提供者の作業能力・人数に依存し，その負担は大きい．

2.3 ディレクトリサービスの問題点
– 利用者の操作コスト

ディレクトリサービスでは，登録サイト数が多いカテゴリは，それをより詳しいカテゴリとして，細かく分類していく．これは，前節で述べたように，サービス提供者の負担となる作業であるが，利用者にとっては，より自分の興味に近いカテゴリをたどることができる．このカテゴリの分類の構造は，利用者のイメージに一致しなかったり，階層が深くさらに最下位のカテゴリ数が多くなると，操作コストの増加につながるという問題がある．

利用者は，自分の必要とするカテゴリを選択し，そのカテゴリ内のサイトを見ることになる．このとき，利用者は，紹介文を頼りに，自分の必要とするサイトにアクセスすることになる．しかし，利用者が紹介文だけでは不十分と思った場合には，そのカテゴリ内の登録サイトを順にアクセスするしかない．この問題を解決するために，カテゴリ内の登録サイトは，辞書順や登録順でソートしているディレクトリサービスが多い(表1参照)．しかし，利用者は，紹介文・サイト名以外に頼るものがなく，操作コストの増大につながっている．

表1 主な検索サイトの表示方法
Table.1 WWW search sites examples.

YAHOO! Japan[3]	辞書順
NETPLAZA[4]	新着順,クリック回数順
NTT DIRECTORY[5]	新着順,辞書順,クリック回数順
iNET Guide[6]	登録順
Infoseekチャンネル[7]	登録順,更新日順
100hot.com[9]	アクセス数順
Lycos Top 5%[10]	内容格付け順,デザイン格付け順

2.4 現在の検索表示方法について

ディレクトリサービスは表1に示したように，利用者がより効率的な検索を行えるように，さまざまな検索表示方法を提供している．

例えば，ディレクトリサービスのNTT DIRECTORY[5]では，“新着順”，“辞書順”，“URL順”，“関心度”，“MYブックマーク登録数順”などの10通りから1つ選択することができる．多くのソート項目から選択できるため，利用者には自由度が高いと思われる．しかし，ソートに用いる項目が1つで，異なる視点からの選択を利用者に提供しているだけにすぎない．

一方，検索エンジンのフレッシュアイ[8]では，キーワード適合度で並び替える“良いもの順”と，更新した日付順で並び替える“新しい順”に，利用者が変更できる．“新しい順”の集計対象は，ロボットで収集したものであるため，利用者にとっては有用であると考えられる．しかし，ソートに用いる項目は1つであり，これ以外の項目では並び替えることができない．

検索エンジンのGoogle![11]では，既存の検索エンジンと同様にキーワード適合度に加えて，リンクの相関関係によるページ重要度を加味して検索結果の表示順序を算出している．しかし，利用者がキーワード適合度とページ重要度の重みを制御することはできない．さらに，キーワード適合度は検索エンジン特有の要素でありディレクトリサービスへの適用は困難であると考えられる．

また，ディレクトリサービスのLycos Top 5%[10]では，“Content(内容の格付け評価)”，“Design(デザインの格付け評価)”の主観評価の要素と，この2つの要素を組み合わせた“Overall”の3通りで，並び替えることができる．レビュアーと呼ぶ格付け担当者による主観的な評価からの要素で算出しているため，その正確性は高いと考えられる．しかし，2つの要素を組み合わせた“Overall”では，2つの要素の組合せ方法は公開

されておらず，しかも利用者による組合せを制御することはできない．このため，利用者の望む並び替えと必ずしも一致しない場合には，利用者の操作コストは増大するという問題がある．

3. WWW 検索表示方法の考え方

3.1 目的

2.3節で述べたディレクトリサービスでの問題点を解決するために，登録サイトに対して，サイト評価情報を設定し，その評価順でソートし表示することによって，利用者が容易に並び替えを制御することを考える．これにより，利用者の操作コストの軽減を目指す．さらに，提案する検索表示方法には以下の特徴を有する．

- サイト評価情報は，1つの要素からではなく，複数の要素から算出する．
- 利用者は，複雑なインターフェイスではなく，バランスレシオと呼ぶ，1つのパラメータを変化することで，複数の要素を組み合わせたソートの方法を制御することができる．

3.2 サイトスコア

登録サイトに対して設定するサイト評価情報（サイトスコア）は，自動的に収集・解析できる要素から算出するものとした．本論文では，アクセス度・更新度・内容度の3つの要素を，サイトを評価できる情報として仮定した．この仮定は，利用者がWWWサイトにアクセスする際の代表的な基準として，多くの利用者からアクセスされているサイト，最新の情報を提供しているサイト，内容が充実しているサイトの3点であると想定したことによる．

サイトスコアは，上記の3つの要素ごとに重み付けし，その総和でサイトスコアを算出する．表示する際は，このサイトスコアでソートする．

3.2.1 アクセス度

アクセス度は，WWWサイトの人気を評価できる要素であると考える．WWWのProxyサーバが記録するアクセスログより解析する．WWWのProxyサーバは，クライアント側にあるサーバである．例えば，大学のProxyサーバには，すべての学生・教職員がWWWにアクセスした情報が記録される．企業のProxyサーバでも同じである．企業と教育機関とを比べると，WWWの利用状況・利用目的が異なるはずである．そのため，Proxyサーバのアクセスログには，それを利用するそれぞれのクライアントコミュニティが反映される．偏りのない正確なアクセス解析を行うためには，現在のWWW利用状況に合わせて，さまざまクライアントコミュニティから複数のProxyサーバのアクセスログが必要になる．

3.2.2 更新度

更新度は，WWWサイトの情報鮮度を評価できる要素である．また，更新の頻度を把握でき，サイトの運営状況も評価できる要素であると考える．WWWサイトを巡回してページを収集し，ファイルのタイムスタンプから解析する．サイトのトップページが常に更新されているとは限らないので，トップページからいくつかの階層分のページも収集する必要がある．また，ファイルのタイムスタンプだけを比較するのではなく，更新された量も比較する必要がある．

3.2.3 内容度

内容度は，WWWサイトのコンテンツの充実さを表す要素である．WWWを巡回して，サイト内のすべてのページを収集し，コンテンツ情報（ページ数・ファイルサイズ）を解析する．

3.3 利用者インターフェイス

利用者が，自分の要求に合わせて，要素ごとに重み付けを同時に変更できるインターフェイスを提供する．すなわち，バランスレシオと呼ぶ1個の変数を提供し，この変数により，要素の重み付けを変化させる．したがって，利用者はマウス操作だけで，容易に並び替えの制御を行うことができる．

4. WWW 検索表示方法

4.1 構成

全体の構成を図1に示す．

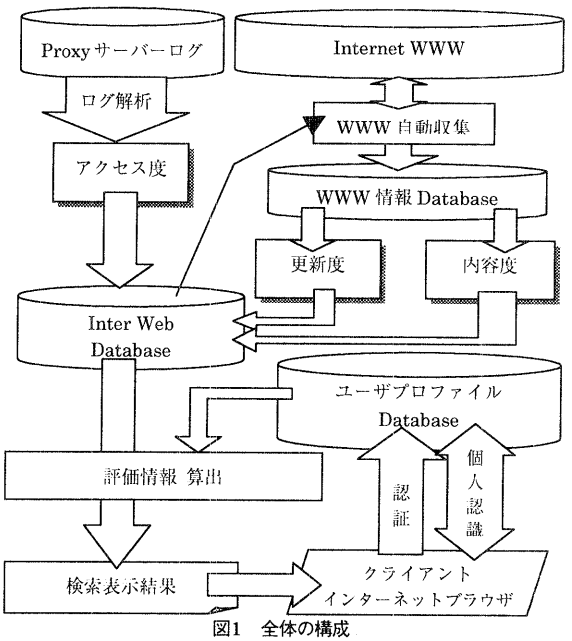


図1 全体の構成

Fig. 1 System structure

収集ロボットは，ディレクトリサービスに登録されているサイト内のすべてのページを収集する．このシステムで扱う“サイト”は，ディレクトリサービスに登録されているURLを1つの単位とし，そのURL以下のページの集合を1つのサイトとした．

複数のプロバイダ・企業や大学内など，ネットワーク上にあるProxyサーバに貯まったログを定期的に転送・収集し，それをもとに各サイト・ページのアクセス数を抽出する．

利用者が設定するバランスレシオは，ユーザプロファイルデータベースで利用者ごとにカテゴリごとに管理され，クライアントブラウザのCookie機能[13]で利用者を識別する．これにより，利用者が以前に設定したパラメータは保存され，ある1つのカテゴリではいつも，自分の要求に近い検索表示にすることができる．

4.2 インターフェイス

提案するWWW検索表示方法で用いて試作したインターフェイスを図2に示す．利用者は，一次元上での点（ブラウザでは，ラジオボタン）を指定することで，3つの要素の重み付けを同時に変更することができる．この一次元上での線をバランスレシオと呼ぶ．

例えば，図2の「人気～新鮮」のバランスレシオ（アクセス度:100～0, 更新度:0～100, 内容度:10～10）では，「新鮮」側にバランスレシオを近づけると，更新度が強調されたサイトスコアになり，登録サイトは更新がよく行われているサイト順に表示される．「人気」側に近づけると，アクセス数が多いサイト順に表示される．

また，利用者は「人気～新鮮」「内容～鮮度」「定番～情報」の3つから，1つを選択することで，バランスレシオのタイプを変更することができる．この変更で，アクセス度・更新度・内容度それぞれの両端の重み付けが変わる．「内容～鮮度」のバランスレシオ（アクセス度:10～10, 更新度:0～100, 内容度:100～0）では，「内容」側ではデータベースサイトなど内容が充実したサイトが，「鮮度」側ではニュースサイトなどの情報鮮度が高いサイトが，各々強調される．「定番～情報」のバランスレシオ（アクセス度:90～20, 更新度:50～50, 内容度:20～90）では，「定番」側では，よく利用されるサイトが，「情報」側では，必要な情報が入手しやすいサイトが，上位に表示される．

4.3 各要素の要素値とサイトスコアの算出方法

提案する検索表示方法のソートに用いるサイトスコアと，アクセス度・更新度・内容度の算出方法について説明する．

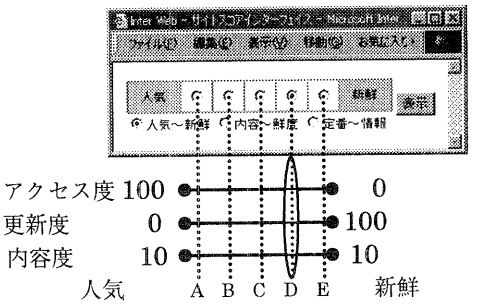


図2 インターフェイスと要素の重み付けの関係
Fig. 2 User Interface and relationship with weight values

4.3.1 アクセス度

1サイトのアクセス総数 A_{access} は，その対象をサイトのトップページに限定すると階層下のページのアクセス数が無視されるため，正確な算出とはいえない．そこで，階層下のページのアクセス数も含めた算出を考える．トップページを0階層目とし，トップページからサイト内リンクされているページを1階層目としていく．

トップページから l 階層目のアクセス数 a_l は，Proxyサーバのアクセスログを解析し，計算した数量である．アクセスログからは，1クライアントが1サイトに対して，多くのアクセスをカウントできる．これは，1クライアントが，1ページのみにアクセスしていることはあり得ず，HTML以外のファイルタイプを含めて，1サイト内の多くのファイルにアクセスしているからである．サイトのアクセス数を正確に計算するためには，アクセスログの日時から判断して，この状況を1カウントとして計算する必要がある．トップページから階層下のページにいくつアクセスしても，階層下の1ページのみにアクセスしても，1カウントと計算される．次式で1サイトのアクセス総数 A_{access} を求める．

$$A_{access} = \sum_{l=0} a_l$$

A_{access} :1サイトのアクセス総数

a_l :トップページからl階層のアクセス数

求めた A_{access} からより，アクセス度 F_{access} を算出する．最大の A_{access} からの割合で求めると，ほとんどのサイトにおいてアクセス度が0に近い値になりうる．そこで，A_{access} の最大限界値を決めておく．これにより，最大限界値以上のアクセス総数になるサイトは，アクセス度が100となる．次式でその最大限界値に基づく割合でアクセス度 F_{access} を求める．

$$F_{access} = \min(A_{access}, A_{\max}) / A_{\max} \times 100$$

F_{access} :アクセス度(0～100)

$A_{\max}$: A_{access} の最大限界値

4.3.2 更新度

更新度は，ページが最後に更新されてからの経過日数と，ページが更新された頻度，更新されたページ数・変更量から，算出する．

サイト内での最後に更新されたページのタイムスタンプを，そのサイトの最終更新日とする．次式で更新されたからの経過日数における度数 L_{update} を求める．

$$L_{update} = \max(d^{L}_{\max} - d_o, 0) / d^{L}_{\max} \times 20$$

L_{update} : 最終更新日における度数

$d^{L}_{\max}$: 最大更新間隔日数

d_i : 最新i番目の更新間隔日数

更新頻度は，更新間隔日数における度数の平均から算出する．次式で，ページの更新頻度における度数 T_{update} を求める．

$$T_{update} = \sum_{i=0}^{n} B_{d_i} \Big/ n \times 60$$

T_{update} : 更新頻度における度数

B_x : 更新間隔日数xに対する度数

d_i : 最新i番目の更新間隔日数

n : 記録された更新間隔日数の個数

変更率 S_{update} は，サイト内の全ファイルサイズにおける，変更部分のファイルサイズの割合から算出される．次式で，更新されたページ数・変更率における度数 M_{update} を求める．

$$M_{update} = \left(p_{update} \Big/ p + S_{update} \right) \times 20$$

M_{update} : 更新されたページ数・変更率における度数

S_{update} : 変更率

p_{update} : 更新されたページ数

p : 全ページ数

以上より，次式で更新度 F_{update} を求める．

F_{update} : 更新度(0〜100)

$$F_{update} = L_{update} + T_{update} + M_{update}$$

4.3.3 内容度

内容度は，サイト内の全ページ数と，サイト内の総ファイルサイズから算出する．コンテンツの充実さを測るには，サービス提供者の人手による客観的な評価から算出するのが理想であるが，大量のサイトについて内容の評価を行うには，負担が大きく，現実的な解決とは言い難い．今回は，効率よく内容の充実さを算出することを重要と考え，サイト内の全ページ数とサイト内の総ファイルサイズで，コンテンツの充実さを近似的に測ることができると仮定した．内容度の算出に，サイト内の全ページ数とサイト内の総ファイルサイズを選んだ理由は，自動的に収集・解析ができ，内容度の値の大小が，コンテンツの充実さを推定することができると考えたためである．

全ページ数における度数 $P_{content}$ は，ページ数が多いほど高くなり，最大限界値以上のページ数があるサイトは，度数が 50 となる．次式で，サイト内の全ページ数における度数 $P_{content}$ を求める．

$$P_{content} = \min(p_{\max}, p) / p_{\max} \times 50$$

$P_{content}$: 全ページ数のおける度数

$p_{\max}$: ページ数最大限界値

p : 全ページ数

総ファイルサイズにおける度数 $S_{content}$ は，総ファイルサイズが大きいほど高くなり，最大限界値以上の総ファイルサイズであるサイトは，度数が 50 となる．次式でサイト内の総ファイルサイズにおける度数 $S_{content}$ を求める．

$$S_{content} = \min(s_{\max}, s) / s_{\max} \times 50$$

$S_{content}$: 総ファイルサイズにおける度数

$s_{\max}$: ファイルサイズの最大限界値

s : 総ファイルサイズ

以上より，次式で内容度 $F_{content}$ を求める．

$F_{content}$: 内容度(0〜100)

$$F_{content} = P_{content} + S_{content}$$

4.3.4 サイトスコア

「アクセス度」「更新度」「内容度」の各要素値から，サイトをソートする際に用いられるサイトスコアを算出する．各要素値に対する重み付けは，バランスレシオによって行う．バランスレシオは，ユーザが設定できる値とする．

「アクセス度」「更新度」「内容度」の各要素値 F_{access}・F_{update}・$F_{content}$ は，0 から 100 の範囲の値をとる．この各要素に，ユーザが設定できる値，バランスレシオを与え，各要素のバランスレシオを R_{access}・R_{update}・$R_{content}$ とし，0 から 100 の範囲で設定できるものとする．次式でサイトスコア F_{site} を算出する．

F_{site} : サイトスコア(0〜100)

$$F_{site} = \frac{F_{access} \times R_{access} + F_{update} \times R_{update} + F_{content} \times R_{content}}{R_{access} + R_{update} + R_{content}}$$

R_{access} : アクセス度に対するバランスレシオ

R_{update} : 更新度に対するバランスレシオ

$R_{content}$: 内容度に対するバランスレシオ

5. 評価

このWWW検索表示方法の効果を見積もるために，バランスレシオを変化させたときのサイトスコア順でのランク変動量と，利用者の要求に対してのランク一

致率とで，評価を行った．

5.1 評価方法

「人気～新鮮」間で，バランスレシオを5段階で変化させて，算出されたサイトスコア順での

・ 変動したサイトの平均ランク変動量と，ランク変動量分布

・ ページビューによるランクとの，ランク一致率

を調査し，評価を行った．

入力する登録サイトとして，インターネットサービスプロバイダのBIGLOBE（以下，プロバイダ）で提供されている代表的なWWWサイトから22個を選んで，調査した．今回の評価では，WWW検索表示方法のインターフェイスの効果を見積もる評価であり，検索後の表示方法の並び替えに関するものであるため，

表2 評価に用いる登録サイト

Table.2 List of sites for evaluations

サイト・URL	アクセス度	更新度	内容度
Begin (http://begin.cplaza.ne.jp/)	31	39	24
CAR GRAPHIC (http://car-graphic.cplaza.ne.jp/)	15	4	28
CYBER PLAZA (http://www.cplaza.ne.jp/)	99	83	84
CYBER PLAZA USA (http://usa.cplaza.ne.jp/)	47	48	79
Career Up! (http://netplaza.biglobe.ne.jp/career/)	30	88	76
Creative Farm (http://www.cplaza.ne.jp/cfarm/)	67	90	68
Gallop ONLINE (http://gallop.cplaza.ne.jp/)	15	46	49
GlobalNatureLife (http://gnl.cplaza.ne.jp/)	7	15	64
Hanako Net (http://hanakonet.biglobe.ne.jp/)	58	48	69
MIDILINK (http://midilink.cplaza.ne.jp/)	15	67	42
MUSICCOLORS (http://music.cplaza.ne.jp/)	43	53	82
Money (http://money.cplaza.ne.jp/)	3	27	12
PERSONAL KINGDOM (http://kingdom.biglobe.ne.jp/)	100	84	61
SoftPlaza (http://softplaza.biglobe.ne.jp/)	97	90	89
いくじーず (http://babyweb.cplaza.ne.jp/)	44	85	36
インターネットファンクラブ (http://ifc.cplaza.ne.jp/)	84	66	73
シネマスクランブル (http://www1.mesh.ne.jp/cinema/)	68	85	56
ワールド ワイド 竹村 (http://takemura.cplaza.ne.jp/)	25	51	68
伊達公子 (http://datekimiko.cplaza.ne.jp/)	9	2	28
好っきやねん大阪 (http://www.meshnet.or.jp/osaka/)	50	51	92
電子書店パピレス (http://www.papy.co.jp/)	13	43	43
旅Web (http://www3.mesh.ne.jp/travel/)	59	64	92

20程度のサイトで十分であると判断した．アクセス度は，1998年8月から3ヶ月間のインターネットでの不特定な利用者からのアクセスによる，プロバイダから提供されたアクセスデータを算出対象とした．更新度・内容度は，同じく3ヶ月間，登録サイトのWWW情報を収集し，算出対象とした．評価に用いた登録サイトのデータ（サイト名・URL・アクセス度・更新度・内容度）を表2に示す．

5.2 サイトスコア順で並び替えた検索表示結果

「人気～新鮮」間でバランスレシオを5段階（図2でのバランスレシオ位置を，左からA,B,C,D,Eとする）で変化させときの，各要素での重み付けを図4に示す．

各要素の重み付けの総和から算出されたサイトスコア順で並び替えた検索表示結果を図3に示す．図3は，バランスレシオを5段階で変化させたとき，22個すべ

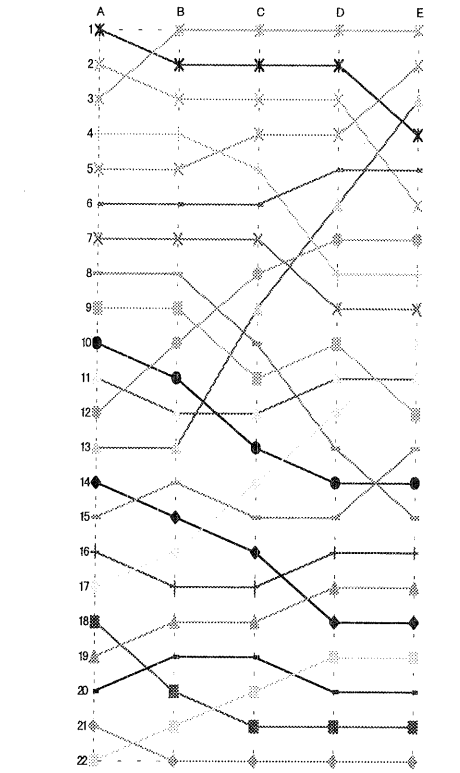
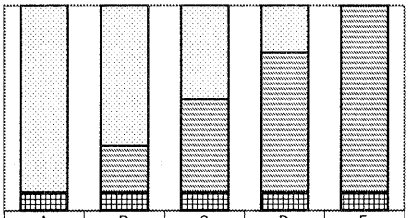


図3 22個のサイトのランク変動

Fig. 3 Ranking fluctuation of 22 sites.



	A	B	C	D	E
□アクセス度	100	75	50	25	0
▨更新度	0	25	50	75	100
▦内容度	10	10	10	10	10

図4 各要素での重み付け

Fig. 4 Weighted values for calculation of balance ratios.

てのサイトのランク変動をグラフ化している．図3に示すように，各サイトのアクセス度，更新度，内容度の値に応じてランクが変動することを示している．以下に，図3をより定量的に分析した結果を示す．

5.3 検索表示結果の評価(平均ランク変動量)

図5に，変動したサイトの平均ランク変動量を示す．バランスレシオをBからC（CからBも同様）に変化させたとき，変動したサイトの平均ランク変動量(変動したランク増減量の絶対値の平均)は，0.87であることを表している．どのバランスレシオ位置からの変化でも，ほぼ一定の平均ランク変動量を示している．すなわち，本検索表示方法は，バランスレシオの変化に合わせて，平均的にサイトのランクが変動していることを示している．

5.4 検索表示結果の評価(ランク変動量の分布)

図6に，すべてのバランスレシオ位置からバランスレシオを変化させたときの，ランク変動量の分布を示す．例えば，図6ではバランスレシオをCからD（DからCも同様）に変化させたとき，ランク増減量が1であったサイトは，9個であることを表している．バランスレシオ位置によって，その分布にばらつきが見られるものの，平均的にはランク変動量は単調減少的なグラフを示す．このことから，本検索表示方法は，バランスレシオの変化に合わせて，極端ではない変動をすることが明らかとなり，利用者による微妙な制御が可能であることを示している．

5.5 検索表示結果の評価(ランク一致率)

図7に，ページビューによるランクとのランク一致率を示す．ページビューによるランクは，全ページにおける総アクセス数順（プロバイダから提供されたアクセスデータ）で，登録サイトをランク付けした．このランクと，サイトスコア順でのランクとの，一致度合いを表すものとしてランク一致率を定義した．2つのランク間で，それぞれのサイトの順位が，すべて一致している場合は，ランク一致率は100%となる．図7に示すように，AからEへ変化させると，ランク一致

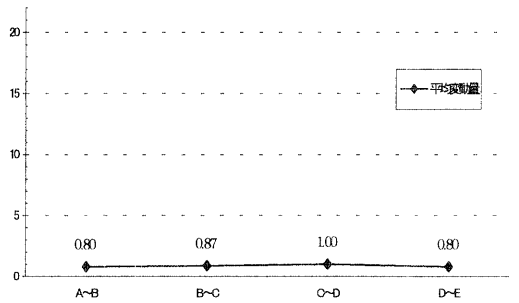


図5 変動したサイトの平均ランク変動量

Fig. 5 the average fluctuation quantity of fluctuated sites.

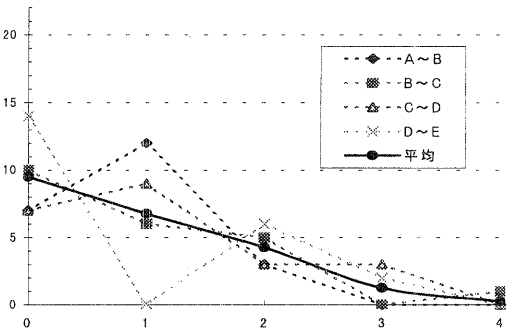


図6 サイトのランク変動量分布

Fig. 6 Fluctuation quantity distribution of sites.

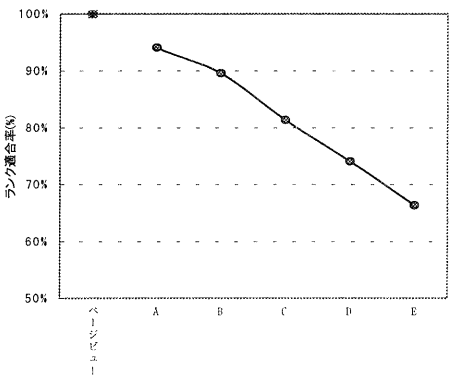


図7 ページビューによるランクとのランク一致率

Fig. 7 Ranking conformity rate between page-view access ranking and sitescore sorted ranking.

率が減少していき，サイトスコアにおいて更新度が強調されていく．すなわち，アクセス数が多いサイトより，更新頻度が高いサイトが上位になるランクとなったことを示している．なお，バランスレシオ位置がEのときでも，高いランク一致率を示しているのは，頻繁にサイトが更新されていると，アクセス数に反映してくると考えられる．また，Aのとき，ランク一致率が100%にならないのは，サイトスコアに内容度が考慮されていることを示している．バランスレシオが「人気」側に近いほど，ページビューによるランクとのランク一致率は高くなっており，利用者がバランスレシ

オを制御することで，要求通りの検索表示となることが分かる．すなわち，利用者がバランスレシオを制御することで，サイトがソートされ表示できる方法を実現することができた．なお，本検索表示方法では，非常に簡単な計算式で算出できるため，実際のディレクトリサービスに適用し1つのカテゴリでの登録サイト数が増えても，十分に対応できるものと考えられる．

6. おわりに

現在のディレクトリサービスの問題点をふまえて，サイトの評価情報を用いたサイトスコアに基づいてソートする検索表示方法を提案した．バランスレシオという1個の変数を用いるだけで，利用者が容易に並び替えの制御をできることが分かった．

ディレクトリサービスを改良する形で，アクセスデータやWWW情報といったデータが主導する新しいWWWガイドとして，より利便性の高いサービスが実現できると考えられる．このようなサービスは，単にWWWの検索という利用方法だけではなく，WWW上でのマーケティング的なビジネスへの利用などでも活用できると考えられる．

アクセスデータは，利用者自身が作り出していくものであり，WWW情報は，サイト運営者が作り出していくものである．こういったデータ主導の検索サービスが機能するようになれば，利用者・サイト運営者がインターネットを，より活性化することにもつながると考えられる．

今後の課題としては，以下のことが上げられる．

- アクセスデータ・WWW情報の収集

  データの収集での技術的な面においては，さらに検討する必要がある．収集したデータの正当性や，データの利用には問題がないのか，検討する必要があり，データ収集を行えるしっかりとした環境づくが必要であると考えられる．
- サイトスコアの算出対象となる要素

  本論文では，アクセス度・更新度・内容度からサイトスコアを算出した．これらの要素はWWW利用者がWWWを主観的に評価している要素と経験的に考えたこと，また，自動的に収集し客観的に分析できる要素である，という点から選んだ．提案した検索表示方法は，この3つの要素で十分であることを，利用者による検索表示結果の主観評価により確認したいと考えている．また，この3つの要素以外にも，レビュアー・利用者が主観評価[10]した要素や，リンクの相関関係からの人気度合い[11]を表す要素など，評価できる要素が考えられるが，これらの要素を追加する必要性，効果を検討していきたい．
- 各要素値・サイトスコアの算出方法

  利用者にサービスとして提供して，その利用状況から，利用者の要求にあった要素値・サイトスコアの算出などを，さらに検討する必要がある．

謝辞

本研究の一部は文部省科学研究費補助金 基盤研究(C)(2)(課題番号10680414)の援助を受けている．また，本研究を進めるにあたりデータの提供を頂いたNEC BIGLOBEパーソナル販売本部に深謝します．さらに，多数の有益なコメントを頂いた本論文誌査読者に感謝致します．

参考文献

1) NEC BIGLOBE CYBERPLAZA, http://www.cplaza.ne.jp/, NEC
2) Sergey Brin, Rajeev Motani, Lawrence Page, Terry Winograd:What can you do with a Web in your Pocket?, Data Engineering Vol.21 No.2 1998, IEEE
3) YAHOO! Japan, http://www.yahoo.co.jp/, 株式会社 ヤフー
4) NETPLAZA, http://netplaza.biglobe.or.jp/, NEC
5) NTT DIRECTORY, http://navi.ntt.co.jp/, NTT
6) iNET Guide, http://www.inetg.com/, CyberSpace Japan, Inc.
7) infoseek Japan, http://www.infoseek.co.jp/, Infoseek Corporation.
8) フレッシュアイ, http://www.fresheye.com/, 東芝
9) 100hot.com, http://www.100hot.com/, Web21社
10) Lycos Top 5%, http://point.lycos.com/categories/, Lycos Inc.
11) Google!, http://www.google.com/, Google Inc.
12) 林良彦,小橋喜嗣: WWW上の検索サービスの技術動向, 情報処理 Vol.39 No.9 Sep. 1998
13) 佐藤正博:クッキーの利用を考える, OPENDESIGN 1999 2 No.30, CQ出版社

(平成11年3月20日受付)
(平成11年6月27日採録)

(担当編集委員 河野浩之)

灰原　清太郎

平成11年3月立命館大学理工学部情報学科卒業．現在，同大学大学院理工学研究科情報システム専攻博士前期課程に在学中．情報検索，エージェント技術，データベースの研究に従事．

川越　恭二(正会員)

昭和50年3月大阪大学工学部電子工学科卒業，昭和52年3月同大学大学院工学研究科電子工学専攻前期課程修了．同年4月日本電気(株)入社．平成9年4月より立命館大学理工学部情報学科教授．情報システム，ネットワークサービス，データベースの研究に従事．博士(工学)．情報処理学会，IEEE等各会員．