

# インターネットにおける サイトランキングの推定

\*Dr. キッチュ

## はじめに

インターネットのホームページを開設している者にとって、最大の関心事はアクセス数の向上である。(図1) アクセス数を増やすために、これまで文献1や、文献2等のサイトで、そのノウハウが議論されてきた。

さて、このように努力して得られたマイ・ホームページは、この膨大なネット世界の中で、どの程度の順位に位置付けられるのだろうか？

今回、Webにおける通信量のデータを、文献3における調査結果から知ることができた。

この情報から、Web上のサイト分布を推定し、日々のカウント数から自分のサイトが世界的にどの程度のランクにあるのか計算により推定したので報告する。



図1 Webページにおけるアクセスカウンタの例

## 前提条件の検討

前出の報告書によれば、1998年8月の段階で、以下の事実が判明している。

- Fact1 現在のWebは3テラバイトに達している
- Fact2 8ヶ月置きに2倍になる
- Fact3 2000万のコンテンツエリア(=サイト?)が存在する
- Fact4 Webトラフィックの90%は10万のホストマシンに広がっている
- Fact5 全トラフィックの50%は900のトップWebサイトに行く

2000万(!)ものコンテンツが存在するという事実だけでも、現在のWebが文字通り「爆発的に」拡大している事実が実感できる。

サイトランキングの検討は、この事実に基づいたサイトのhit数分布を推定する事により可能となる。

## 計算手法について

サイトの分布形状は、多分にErfc(補誤差関数)になるものと思われるが、今回は単純化のため、次

\* 超越科学研究所・ワークスキッチュ  
マッドサイエンス学会正会員  
Laboratory of Hyper-Science  
Tokyo JAPAN  
dr\_kitsch@muf.biglobe.ne.jp

の仮定を用いた。

仮定1 水準Nと水準N+1のサイトの日当たりhit数の比率はNである。

仮定2 水準Nと水準N+1のサイト数の比率はNである。

ここで、水準とは、日当たりhit数がある範囲にあるサイトの集合である。律儀に全順位を計算することは、メモリー不足や見通しの悪化を招くので、今回はこのようなモデルを作り実施した。

計算上では、水準は100水準である。上位3水準程度はそのままサイト順位にあてはまり、その後同じ順位に同着のサイトが増加していくイメージである。

NとNの決定には、先の報告書にある数値を適用した。すなわち、

$$(全水準のサイト数)=2000万 \quad (1)$$

$$(上位900サイトにおけるHit数)$$

$$= (全サイトにおけるhit数)*0.5 \quad (2)$$

$$(上位10万のサイトにおけるhit数)$$

$$= (全サイトにおけるhit数)*0.9 \quad (3)$$

である。このグラフを図2の'98/08に示す。

## 順位の計算

日当たりhit数におけるランキングのグラフが求められたので、あとは各自のサイトにおける日当たりhit数を当てはめれば、自分のサイトの全世界的ランキングが判る。ちなみに、先の調査結果は、アダルトサイトも含めた調査であるので、アダルトサイトだからもっと上位に違いないということは有り得ない。

例として、つきなみCOMICSサイトにおけるランキングを計算してみよう。98/08時点での日当たりhit数はおよそ20~30hit/dayだったので、グラフより47万番目であることが判る。

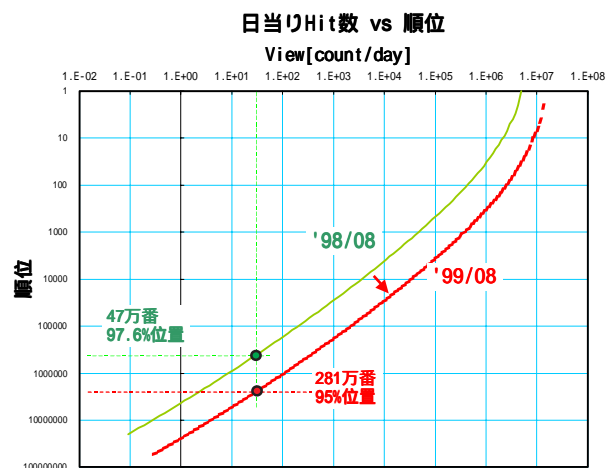


図2 Webページ日当たりHit数に対するランキング順位

たとえば、このhit数が1桁ランクアップすると、同じくグラフより7万8千番目と、ほぼ1桁の躍進となることもわかる。

このことから、いかにhit数を増加させるかが重要であることが理解されるであろう。

## 情報価値の減価償却の効果

ここまでの議論は、あくまでも1998年8月時点におけるサイトランキングの推定値に過ぎない。

先程の調査結果にあるとおり、驚くべきことに Web ページは 8 ヶ月で 2 倍のスピードで増殖しているのだ。

半導体業界で有名なムーアの法則ですら 2 倍の規模になるのに 18 ヶ月かかっているのだから、1 年あたりの乗率で 1.8 倍も増加係数が高い。逆に言えば、Web ページの時間は半導体時間の 2.3 倍のスピードがあるという事になる。従って、この 1996 ~ 1999 年の 3 年間は、パソコン進化での 7 年にも相当するという事だ。

すでに文献 4 で報告している様に、PC のソフトウェア寿命は約 3 年と見積られている。これを Web 時間に換算すると、一般的な Web ページ寿命は

$$3/2.3 = 1.3 \text{年} \quad 16 \text{ヶ月} \quad (4)$$

と見積れる。

したがって、この 16 ヶ月で情報価値は償却してしまったと見なさねばならない。

このことをふまえて、1999 年 8 月における Web サイトの分布を図 2 の '99/08 に示す。

カウント数の伸びがなかったため、同じ日当りのカウント数であっても、順位は 281 万番目になってしまっている。

もちろん、この数値が大きくなったのは、Web ページ数が増大しているためもある。

このため、全 web ページに対する順位割合も計算した。(図 2 の % 値) この場合、トップが 100%、最下位が 0% である。この相対的位置付けからみても、つきなみ COMICS は 97% から 95% 台に降下してしまっただけが見て取れる。

この結論も、いかに早期にスタートして、そこで知名度をあげるかが重要であることを物語っている。最もアクセス向上に効果ある対策、それは一秒でも早く Web サイトを立ち上げることなのだ。完成度というファクターより、時刻のファクターが厳しい世界なのである。

### アクセス順位増加対策の考察

図 2 のグラフが得られたので、対策を具体的に考えてみよう。

Web サイトが増加することを、文献 5 と同様な考え方で把握すると、それは情報量の増加とみなせる。ここでの結論では、情報量は

$$Ro \cdot \exp(r \cdot t) \quad (5)$$

で変化することになっていた。

同様に、Web ページ増加指数 = 情報増加指数と考えると、1 日あたりの情報増加量  $r$  は、

$$r = 2^{1/(30.5 \cdot 8)} = 1.0028 \quad (6)$$

である。これは、1 週間で約 2% の増分に相当する。

従って、Web 運営者は自分の保有するコンテンツのうち 2% 以上に相当する部分を増加・置換することにより Web 減価償却を乗切ることができる。

### 新しい Web 運営戦略

以上の考察から、有効な戦略も考察してみよう。情報増加量を 2% にキープするために、Web サイトをどんどん拡大していったらどうなるであろうか。

この場合、サイトの全情報量も増大するので、追加が必要なコンテンツの分量も、8 ヶ月で 2 倍必

要になってしまうのだ。

一般的に個人でサービスできる情報量には限界があると考えられる。よって、このような拡大戦略は必ず破綻をまねく。

このあたりも、生物の進化モデルに近似した考え方が適用できる。すなわち、2 通りの戦略があげられる。

戦略 1 サイトの上限を決めて、情報を置換していく。(ゾウリムシの核交換戦略)

戦略 2 サイトの大きさが一定値を越えたところで、2 つのサイトに分裂する。(細胞分裂戦略)

個人が管理運営するのに適しているのは、明らかに戦略 1 の方であろう。

ここで明確な結論が得られる(図 3)。

### Web サイトの生き残り戦略とは、古い情報を交換し捨てることにあり

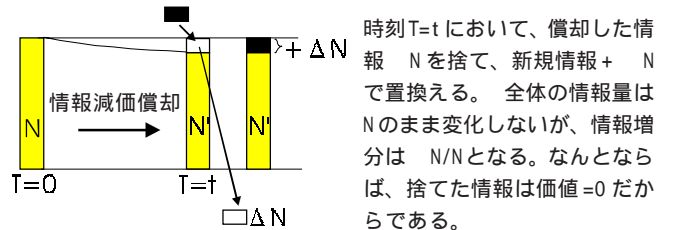
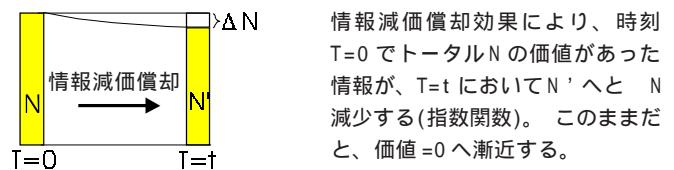


図 3 効果的なサイト運営戦略

### まとめ

インターネットにおけるサイトランキングを考察し、サイト分布を仮定することで順位推定が可能になった。

この分布をつきなみ COMICS の Hit カウントに適用し、ランキングを計算した結果、98/08 において 2000 万中 47 万番、99/08 現在において 5657 万中 281 万番であった。

さらに、情報減価償却という概念にてサイト運営を考察し、毎週 2% 以上の情報交換(追加ではない)が最も有効であるとの見通しを得た。

さっそく、この結果をつきなみ COMICS サイト運営に応用していきたい。

### 参考文献

- 1) アクセス向上委員会  
<http://www.access.or.jp/>
- 2) ズバリ!! 見てもらえるホームページの作り方  
[http://www.asahi-net.or.jp/~LE9S-ICKW/how\\_.html](http://www.asahi-net.or.jp/~LE9S-ICKW/how_.html)
- 3) ALEXA Internet  
<http://www.alexa.com/company/inthenews/webfacts.html>
- 4) T.M.S.R. Vol.12 「98 アーキテクチャの危機と展望」
- 5) T.M.S.R. Vol.15 「情報から見た、人類進化の量的把握」