全世界の Web サーバの地理的位置・バックリンク数の解析

Analyzing Geographical Location and Number of Back-Links of Web Servers All Over the World

平手 勇宇♥ 片瀬 弘晶♦ 山名 早人♠
Yu HIRATE Hiroaki KATASE
Hayato YAMANA

Web サーバからは，膨大な情報が発信され続けており，我々の調査では 2006 年 10 月時点で世界中に 537 億のページが存在すると予測している．我々はこれまでに e-Society プロジェクトで収集した約 107 億ページのトップレベルドメイン分布・言語分布の調査，および約 4,700 万 Web サーバの地理的設置位置の特定を行ってきた．本稿では，これまでの解析の発展として，トップレベルドメイン毎の Web サーバの地理的設置位置分布，Virtual Host 数の地理的分布，および Web サーバ単位でのバックリンク数分布の解析を行った．面積の小さな島国のトップレベルドメインに属する Web サーバは，当該国の外部に多く設置されていること，1ホストあたりの Virtual Host 数が多い国は，中南米・東欧諸国であること，そしてバックリンク数と Web サーバ数は Power-Low に従っていることを確認した．

According to our investigation result in Oct. 2006, the number of Web pages all over the world is estimated 53.7 billion. We have crawled about 10.7 billion Web pages from about 47 million Web servers from all over the world. Then we have investigated TLD distribution and Language Distribution of Web pages, and geographical locations of Web servers based on our dataset. In this paper, as one of our Web statics investigation series, we report three kinds of analysis, geographical distribution of Web servers by Top Level Domain, geographical distribution of the number of virtual hosts per one Web server, and the number of back links, i.e. the value of in-degree, per one Web server. Our results show (1) Web servers belonging to Top Level Domain of small island countries are placed outside the countries with high probabilities, (2) Web servers located in Latin America and East Europe have a large number of virtual hosts, and (3) the distribution between the value of in-degree and the number of Web servers follow the power low.

1. はじめに

近年，Webサーバから発信される情報量が膨大になり，2006年10月時点で世界中に537億のページが存在するという調査結果が報告されている[1]．このような大量のWebページから有用な知識を抽出することを目的とし，リンク解析をはじめとするWebマイニングに関する研究[2][3]が幅広く行われている．しかし，統計的に意味のある解析を行うためには，Webについての各種統計的な分布を知ることが欠かせない．

実際のWebデータを対象とした解析として，著者らのメンバーである加藤らは，2006年にe-Societyプロジェクト[4]で収集を行ったcomとjpドメインを主対象とした約30億ページデータをもとに，TLD分布，言語分布，およびリンク解析を実施した[5]．しかし，解析対象のWebページは，com，jpに偏っており，Web全体を網羅した解析にはなっていない．

そこで我々は，2007年度より，同じくe-Societyプロジェクト[4]で収集した約107億ページを対象として解析を開始した．これまでに，WebページのTLD分布，言語分布，Webサーバの設置位置分布についての解析を行った[6][7][8]．その結果，「com」に属するWeb サーバが全Webサーバの37.8%，英語で記述されたページが全Web ページの42.6%，米国に設置されているWebサーバは，全Webサーバの48.4%を占め，現在のWebが「com」，「英語」，「米国」に偏っていることを示してきた．

本稿では，[6][7][8]の解析の発展として，以下に示す３つの解析結果を報告する．

- トップレベルドメイン（以下，TLDとする）別に，TLDを管理している国の国内に設置されているWebサーバ数（以下，国内設置Webサーバ数とする），国外に設置されているWebサーバ数（以下，国外設置Webサーバ数とする）の分布の計測
- 設置国別に，Webサーバ数とIPアドレス数を比較することにより，Virtual Host数の多い国・地域の特定
- Webサーバ単位のバックリンク数(=In-degree数)の分布

本稿では，以下，次のような構成を取る．第2節において，Webサーバの地理的設置位置に関する関連研究，およびWebページのIn-degree数分布に関する研究を示す．第3節において，解析対象のデータセットと解析手法について示す．第4節において解析結果を示し，第5節にてまとめを行う．

2. 関連研究

2.1 Web サーバ地理的位置特定に関する研究

中平らは，2006年にアフリカの国別ドメイン(=ccTLD)に属する1,600のWebサーバに対し，当該Webサーバの設置位置の特定を行っている[9]．[9]では，自国内に設置されていたホストは約20%であり，残りの約80%は主に欧州，北米に設置されていたという報告がされている．また，2007年に近藤らがIPアドレス100万件を対象に，Webサーバの設置位置を特定した結果，約52.8%のWebサーバが米国内に設置されていたという結果を報告している[10]．

このように，Webサーバの設置位置を特定する研究は行われているが，調査対象のWebサーバ数が少ないことや，地域的に偏っている問題点をもつ．そこで我々は，e-Societyプロジェクト[4]で収集した約107億ページのWebデータを用いて，全世界のWebサーバの地理的設置位置の特定を行った[8]．解析対象のWebサーバ数は，約4767万サーバである．その結果，図 1に示す通り全世界のWebサーバは，北米，欧州，アジアに集中しており，全Webサーバのうち約95.5%が上記3つ

♥ 正会員　早稲田大学メディアネットワークセンター
hirate@yama.info.waseda.ac.jp
♦ 学生会員　早稲田大学基幹理工学研究科
katase@yama.info.waseda.ac.jp
♠ 正会員　早稲田大学理工学術院／国立情報学研究所
yamana@waseda.jp

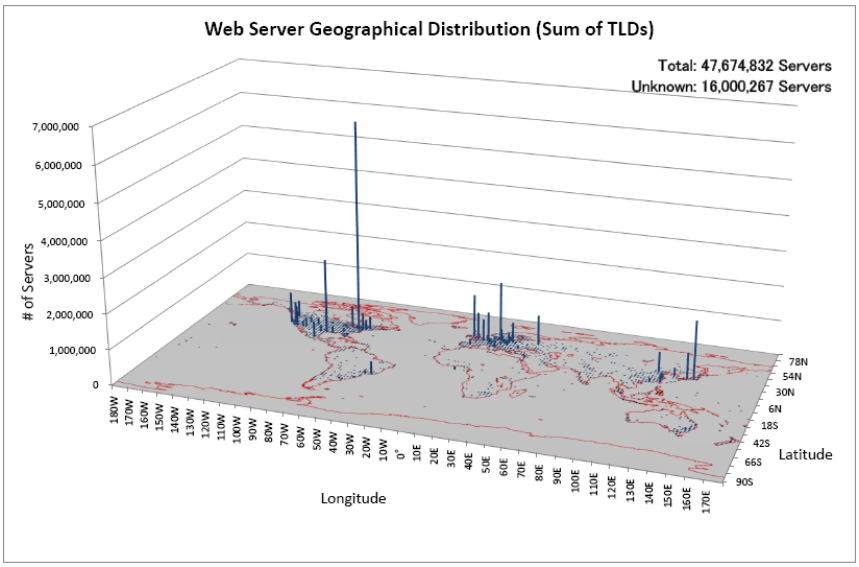

図 1　全世界の Web サーバの設置位置の分布

表 1 Web サーバの設置数上位 20 カ国

設置場所	Webサーバ数	割合
United States	15,325,293	48.8%
Japan	2,437,179	7.69%
United Kingdom	2,060,505	6.51%
Germany	1,937,323	6.12%
Poland	1,048,544	3.31%
Russian	889,628	2.81%
China	830,577	2.62%
Netherland	827,878	2.61%
Canada	780,223	2.46%
France	730,039	2.30%
Brazil	500,162	1.58%
Italy	469,912	1.48%
Mexico	455,267	1.44%
Denmark	335,511	1.06%
Switzerland	235,851	0.74%
South Korea	200,617	0.70%
Taiwan	217,768	0.69%
Sweden	216,908	0.69%
Czech Republic	212,351	0.67%
Spain	198,380	0.63%

の地域に設置されている結果となった．ここで**図 1**右上の「Total」の数字は，我々が収集を行った全てのWebサーバ数を示しており，「Unknown」の数字は，設置場所を特定できなかったWebサーバ数を示す．なお，設置場所が特定できなかった理由は，ドメインからIPアドレス解決できなかった，もしくはIP-経度・緯度変換テーブル[11]でのエントリがなかった理由が挙げられる．さらに，国別に詳しく見ていくと**表 1**に示す通り，全世界のWebサーバのうち約48.4%のWebサーバが米国内に設置されている結果となった．

これに加え[8]では，jpドメイン・tvドメインに属するWebサーバの設置位置を示し，国内設置Webサーバ数と国外設置Webサーバ数の関係は，TLDによって違うことを示した．これらを踏まえ，本稿ではその発展として全TLDの国内設置Webサーバ数・国外設置Webサーバ数の関係の解析を行う．

2.2　In-degree 値の分布に関する研究

これまでに，Webページのバックリンク数(=In-degree値)の分布に関する様々な研究が行われている．すべての研究において，WebページのIn-degree値と，Webページ数の分布はPower-lowに従っていると報告されている．一般に，Power-lowに従っている場合，In-degree値xとIn-degree値に対応するWebページ数yの関係において，式(1)のような関係が成立する．

$$y = ax^{-\gamma} \qquad (1)$$

ここで，aは定数を表し，指数γはPower-Lowの強さを表す．

Kumarらは1999年にAlexaで収集した約20万ページのWebページデータを対象にγの値を計算した結果，おおよそ2であると結論づけている[12]．また，Barabashiらは同年に，約80万ページ に対して計測した結果，γ=2.1[13]，Borderらは2000年に，AltaVistaの約20.3万ページのデータに対して適用させた結果，γ=2.1[14]，Boldiらは2002年にアフリカの約200万ページに対して解析したところ，γ=1.92[15]，Liuらは2005年に，Peking University Sky net Search Engineで収集した約140万ページを基に，γ=2.05であった[16]とそれぞれ報告している．以上をまとめると，**表 1**に示す通りとなる．

これらの既存研究は，対象とするWebの年代が古いという問題点や，対象Webが地域的に偏っているという問題点が存在する．これに対して，本稿では2006年7月までに収集した全世界の107億ページデータを対象に，γの値を示す．

表 2　In-degree に関する既存研究

著者	年	ノード数	γ の値
Kumarら[12]	1999	Alexa で収集した 20 万ページ	約 2
Barabasiら[13]	1999	80 万ページ	2.1
Broderら[14]	2000	Altavista で収集した 20.3 万ページ	2.1
Boldiら[15]	2002	アフリカの Web ページ 200 万ページ	1.92
Liuら[16]	2005	Peking Univ. Sky Net Search Engine で収集した 140 万ページ	2.05

3. 解析対象データ・解析手法

3.1　データセット

解析対象のWebページは，e-Societyプロジェクト[4]によって収集したWebページである．e-Societyプロジェクトでは，2004年1月からWebページの収集を継続しており，2004年1月～2006年7月に合計で14,456,201,906ページの収集を行った．Webページの収集にあたっては，2004年1月の収集開始時に，著者らが保有していたcom，org，edu，net，uk，jp，us，ca，atの9つのTLDから合計600万のWebサーバリストを起点として設定した．起点からは最大15ホップ先までを収集し，新規に発見したWebサーバは起点として随時追加を行った．なお，Webサーバに与える負荷を軽減するため，1Webサーバあたりの最大収集ページ数を5万ページとすると共に，CGIページを発見した場合は，当該CGIページからのホップ数を1に制限した．

2006年7月時点において，世界中で発見したWebサーバ数は約13,468万台であり，内8,116万台の収集を完了した．なお，8,116万台の内，実際に収集できたのは5,548万台であり，2,568万台（収集済サーバの約32%）は既にIPアドレスが存在しない等の理由からアクセスができなかった．また，256万台（収集済Webサーバの3%）については，/robots.txtによりクローラによるアクセスが禁止されていた．残り約5,000万台については未収集である．

2006年7月時点におけるNetcraft社[17]が発表している統計情報によれば，全世界のWebサーバ数 は約8,761万台であり，我々のクローラは，Netcraft社が把握しているWebサーバ数の

約1.5倍のWebサーバを発見することに成功している．以下では，2006年7月までに収集された約144.5億Webページの中から47,674,832台中の10,696,996,553ページ1を対象として解析を行った．

3.2 Web サーバの設置国，設置位置の特定

解析対象の約4,700万台のWebサーバに対し，次に示す手順を踏むことでWebサーバの設置位置の特定を行った．

1. Webサーバ名をDNS参照することによって，当該WebサーバのIPアドレスを特定
2. IP2Location社が提供しているIP-経度・緯度変換テーブル[11]を利用して，当該Webサーバが設置されている設置国，および経度・緯度情報を取得
3. 全世界を，経度2度・緯度2度四方セル2に分割を行い，あるWebサーバがどのセルに設置されているのかの特定

その結果，約4,700万台のうち，31,674,565台のWebサーバの設置国・設置位置の特定を行うことができた．

3.3 IP アドレス数の地理的分布

Webサーバとして利用されているIPアドレス（以下，特に断りがない限りIPアドレスと表現する）の地理的分布の調査を行った．本解析の目的は，同一IPアドレスで運用されているWebサーバ数（Virtual Host数と同義）についての傾向を見出すことにある．

1.2と同様，IP2Location社のIP-経度・緯度変換テーブル[11]を利用し，設置国別及び経度2度・緯度2度四方セル別にIPアドレス数の数えあげを実施した．

3.4 バックリンク数の計算

約107億ページのバックリンクの算出を短時間で行うことは困難である．そこで，本稿ではWebサーバ単位でのバックリンク数の算出を行った．つまり，リンク先のWebサーバが，リンク元のWebサーバと違ったリンクのみを，バックリンク数計算対象のリンクとした．

4. 解析結果

4.1 TLD 毎の Web サーバの設置位置分布

国別トップレベルドメイン(=ccTLD)毎に，当該ccTLDに属するWebサーバの設置位置の分布をすべて計算した結果を，同一グラフにプロットを行うとの**図 2**のグラフが生成できる．**図 2**における1つのプロットは，1つのccTLDの値に対応する．

図 2において，グラフの左上にプロットされればされるほど，国内設置Webサーバ数に比べ国外Webサーバ数が多いccTLDであり，右下にプロットされればされるほど，国内設置Webサーバ数に比べ国外設置Webサーバ数が少ないccTLDであることを意味する．ここで**図 2**を見ると，以下に示す3つの傾向を見ることができる．

- 国外Webサーバ数の割合が極端に多いccTLDは多数存在する．
- 自国内Webサーバ数の割合が約95%以上と極端に多いccTLDは存在しない．
- 自国内Webサーバ数がある一定以上の値となると，自国内Webサーバ数と国外Webサーバ数に相関関係がある．

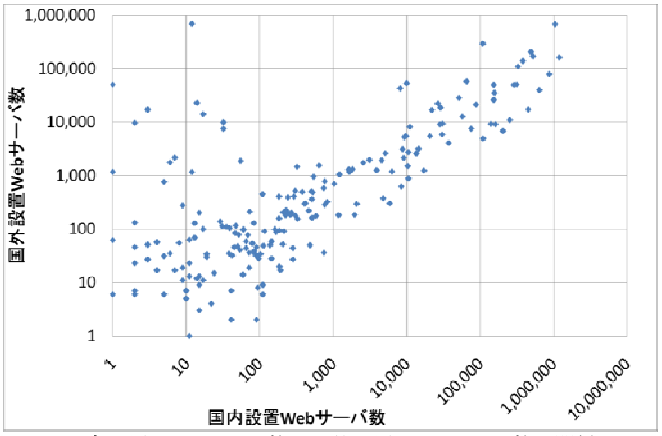

図 2 国内設置 Web サーバ数・国外設置 Web サーバ数の関係

そこで，国内設置サーバ数が60[3]以上であるccTLDに対象を限定し，国内設置Webサーバ数と国外設置Webサーバ数の累乗近似式の算出を行った．その結果，自国内Webサーバ数と国外Webサーバ数の関係は，式(2)のような関係で近似できる．

$$y = 1.076x^{0.863} \quad (2)$$

ここで，yは国外設置Webサーバ数を指し，xは国内設置Webサーバ数を指す．指数部が0.863であることから，国内設置Webサーバ数が多いccTLDほど，国外設置Webサーバ数の割合が小さくなる傾向がある．

さらに，国内設置Webサーバ数に対する国外設置Webサーバ数の標準範囲を，上記近似式を用いて定義を試みる．つまり，標準範囲を逸脱したccTLDは，国内設置Webサーバ数に対して，国外設置Webサーバ数が多すぎる，または少なすぎるといった理由で，アンバランスなccTLDとして検出する．上記近似式を基準値とし，自国内Webサーバ数に対する国外Webサーバ数の標準範囲を仮に基準値の1/10倍以上，10倍以下と設定すると，**図 3**に示すように，自国内Webサーバ数が60以上のccTLDは，ほとんどが標準範囲内に位置する．そこで，自国内Webサーバ数xに対する国外サーバ数yの標準範囲を式(3)のように定義する．

$$\text{標準範囲：} \quad 0.1076x^{0.863} \le y \le 10.76x^{0.863} \quad (3)$$

この標準範囲外に該当するccTLDが，国内設置Webサーバ数と国外設置Webサーバ数の関係がアンバランスなccTLDであると解釈することができる．

図 3に標準範囲外のccTLDにccTLD名を付与した．自国内サーバ数が多いccTLDにも関わらず，国外Webサーバ数が多いとして標準範囲外となったccTLDは，be（ベルギー），hr（クロアチア），lu（ルクセンブルク）の3ドメインであった．また，国内Webサーバ数が多いとして標準範囲外となったccTLDは，bh（バーレーン），bj（ベナン），bt（ブータン）の3国であった．

また，国内設置Webサーバ数が少ないccTLDの中で，国外設置Webサーバ数が多いとして検出された国々は，おおよそ次に示す二つに分類することができた

- トンガ，ツバル，ココス諸島等のカリブ海．太平洋，インド洋等の大洋の中に存在する島国
- リビア・アフガニスタン・東ティモール等の国内に紛争を抱える国々

[1] Disk 故障等により一部のデータが利用できなくなり，解析にあたっては約 4700 万サーバから収集した約 107 億ページを利用した．

[2] 経度 2 度緯度 2 度四方の領域は，おおよそ関東甲信越地方の大きさと同等である．

[3] 国内設置 Web サーバ数の閾値を，10 から 200 の間で 10 ずつ変化させ，その都度累乗近似式の計算を行った．閾値を 60 とした時，累乗近似式の R-2 乗値が最大となったため，閾値として 60 を選択した．

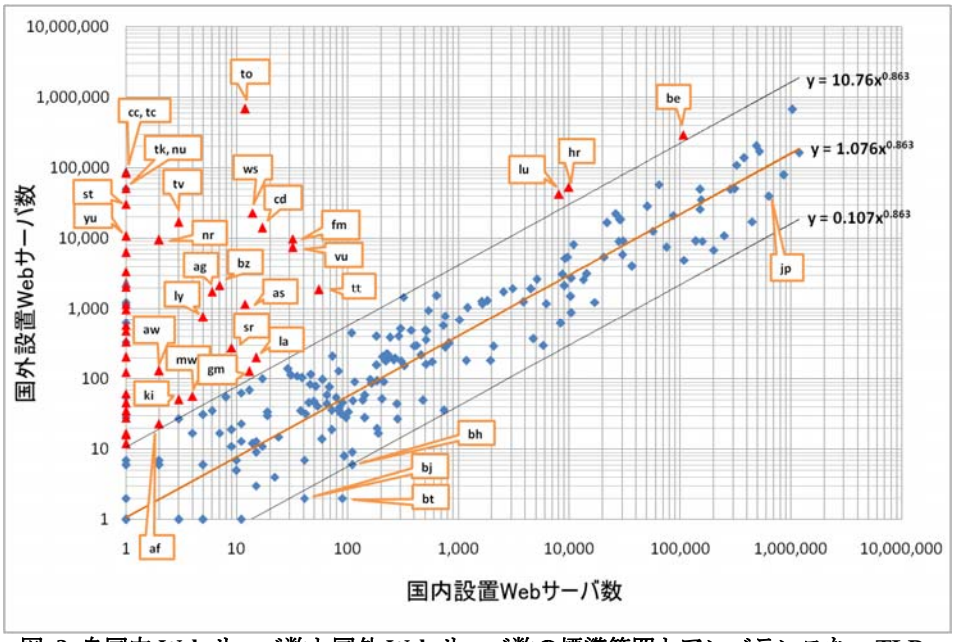

図 3 自国内 Web サーバ数と国外 Web サーバ数の標準範囲とアンバランスな ccTLD

表 3 IP アドレス数上位 20 カ国

設置場所	IPアドレス数	割合
United States	1,020,019	45.69 %
Japan	178,819	8.01%
Germany	178,494	8.00%
United Kingdom	91,300	4.09%
China	79,334	3.55%
Canada	68,570	3.07%
Netherland	52,322	2.34%
Poland	44,502	1.99%
Italy	42,705	1.91%
France	37,011	1.66%
Korea	36,758	1.65%
Russia	34,903	1.56%
Australia	32,120	1.44%
Taiwan	30,230	1.35%
Spain	22,822	1.02%
Switzerland	21,324	0.96%
Sweden	19,510	0.87%
Brazil	18,513	0.83%
Denmark	17,535	0.79%
Czech Republic	12,576	0.56%

表 4 Web サーバ数・IP アドレス数の地域分布

設置地域	Webサーバ数	IPアドレス数	Webサーバ数の割合	IPアドレス数の割合
アジア	3,962,280	360,467	12.51%	16.14%
アフリカ	48,322	8,867	0.15%	0.40%
オセアニア	223,279	38,202	0.71%	1.71%
中東	106,912	15,389	0.34%	0.69%
中南米	1,046,527	46,198	3.30%	2.07%
北米	16,107,027	1,088,797	50.87%	48.76%
欧州	10,169,807	674,513	32.12%	30.21%
その他	1,187	361	0.00%	0.02%

4.2 Web サーバの IP アドレス数分布

約4,700万WebサーバのIPアドレスの経度・緯度情報を算出し，IPアドレスの地理的設置位置の特定を行った．**表 3**に，IPアドレス数の上位20カ国のIPアドレス数およびその割合を示す．**表 1**で示したWebサーバ数上位20カ国と同様，北米，欧州，アジアの3つの地域に属する国が並んでいる．

さらに，**表 4**に世界の地域別のWebサーバ数・IPアドレス数，およびその割合を示す．**表 4**においてWebサーバ数の割合比べて，IPアドレス数の割合が増加した地域は，アジア，アフリカ，オセアニア，中東であり，減少した地域は，中南米，北米，欧州である．この結果から，Virtual Hostを多く立てているWebサーバは，中南米，北米，欧州に多く存在することがうかがえる．

ここで，Webサーバ設置国別にWebサーバ数とIPアドレス数の関係をプロットすると，**図 4**に示すように強い相関があることがわかる．そこで，4.1節と同様に，IPアドレス数に対するWebサーバ数の標準範囲の定義を行い，標準範囲から外れた国をIPアドレス数に対して，Webサーバ数が多すぎる，少なすぎるといった理由で検出を行う．Webサーバ数yとIPアドレス数xの関係の近似式は式(4)に示す通りである．

$$y = 1.11x^{1.23} \tag{4}$$

上記近似式を基準値とし，IPアドレス数に対するWebサーバ数の標準範囲を基準値の1/2倍以上，2倍以下と仮に設定すると，**図 4**に示すように，ほとんどの国が標準範囲内にプロットされる．そこで，IPアドレス数xに対するWebサーバ数yの標準範囲を式(5)のように定義する．

標準範囲： $$2.22x^{1,23} \leq y \leq 0.555x^{1.23} \tag{5}$$

IPアドレスとWebサーバ数の近似式，標準範囲および，標準範囲を逸脱した国の例を**図 4**に示す．

図 4を見ると，IPアドレス数に対してWebサーバ数が多いとして判定された国は，ロシア，ベラルーシ，グルジアの東欧諸国，スロベニア，マケドニアのバルカン半島国々，ブラジル，メキシコの中南米諸国で多数を占められている．これらの国々に設置されているホストは，多数のVirtual Hostを立てている傾向があると結論づけられる．

さらに詳しい分布を実行するために，**図 5**では，経度・緯度2度四方のセル毎に，Webサーバ数基準値と比較して何倍のWebサーバ数が存在するかをグラフ化したものである．すべてのセルに対して，次に示す様な計算を実行した．

1. 当該セル内に設置されているIPアドレス数からWebサーバ数の基準値を計算
2. 実存していたWebサーバ数を基準値で除算

図 5に示す通り，ブラジル国内の西経50度，南緯20度付近のセルが基準値の約130倍のWebサーバ数が設置されており，この付近に設置されているVirtual Host数が異常に多い結果となった．ブラジルの他には，アメリカ内陸部，ポーランド付近，メキシコシティにおいて，基準値と比較して大幅に多いWebサーバ数が検出された．

4.3 In-degree 値の分布

Webサーバのバックリンク数(=In-degree値)の分布状況を**図 6**に示す，**図 6**に示すように，In-degree 値とWebサーバ数の関係にPower-Lowの関係が成立している．ここで，既存研究と同様に，In-degree値xとWebサーバ数yの関係式を示すと式(6)のようになる．

$$y = (1.5e + 7)x^{-1.45} \tag{6}$$

ここで，γの値は1.45となり，既存研究の γ の値よりも小さな値となった．これは，我々の解析がWebページ単位ではなく，Webサーバ単位であることに起因すると考えられる．すなわち，バックリンクをWebサーバ単位で集約を行っているため，

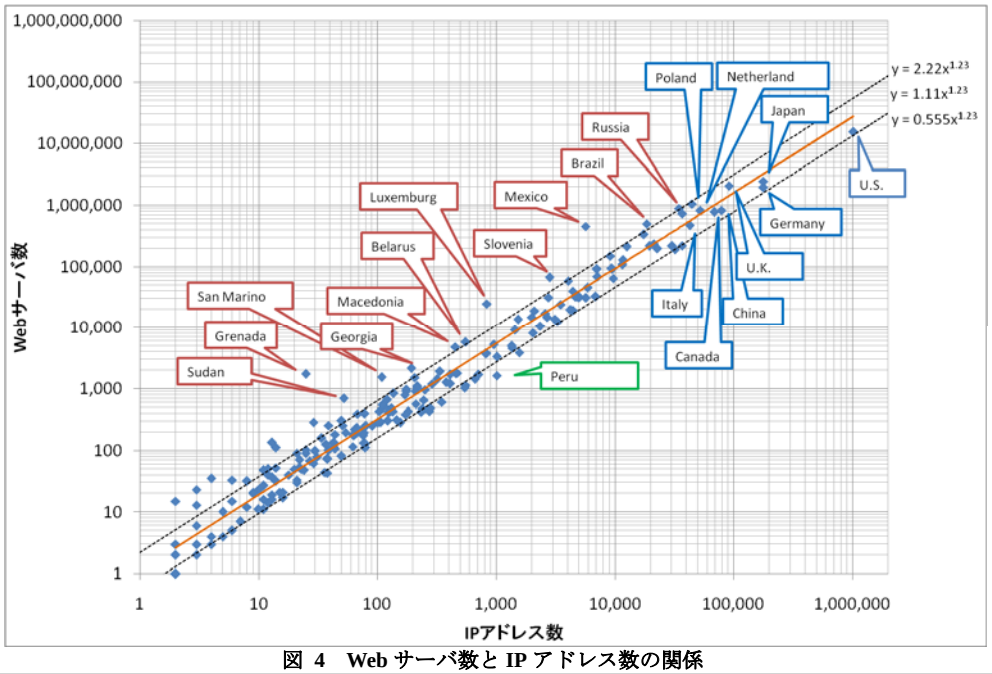

図 4　Web サーバ数と IP アドレス数の関係

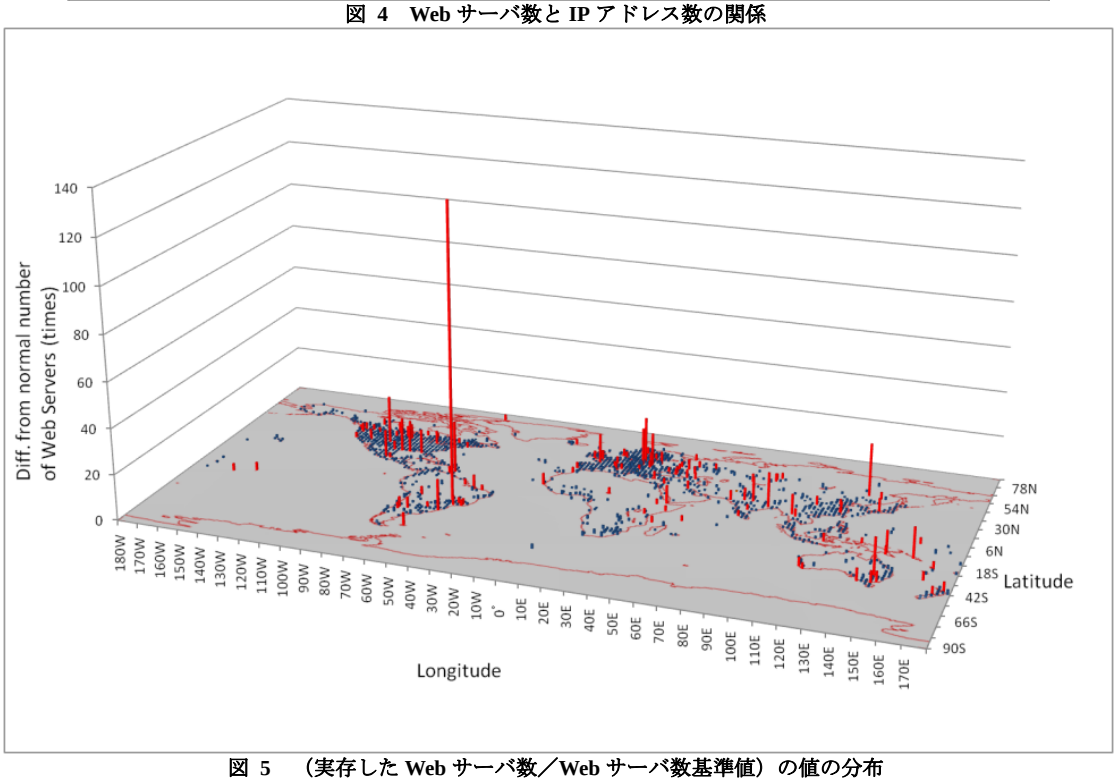

図 5　（実存した Web サーバ数／Web サーバ数基準値）の値の分布
Web サーバ基準値は，当該セル内に存在した IP アドレスから式(4)を利用して算出した．
また，（実存した Web サーバ数／Web サーバ数基準値）の値が大きいセルほど Virtual Host 数が多い地域である．

In-degree値が高いWebサーバが多く存在するためである．

また，図 6の分布において，べき乗分布の関係から一部突出した値が存在している．これらの突出値は，同一バックリンク数(=In-degree数)のWebサーバ数が，通常よりも突出して多いことを意味するので，スパムWebサーバである可能性が高いと考えられる．

図 6中では，In-degree値が520,318のWebサーバが20サーバ存在することが確認できるが，これら20サーバを調査したところ，以下に示す3つを確認した．

- 20Webサーバのトップレベル・セカンドレベルドメインが，同一ドメイン名（toドメイン）
- 20Webサーバすべてが，同一IPアドレス（米国設置）
- 20Webサーバすべてが，同一の520,318Webサーバから被リンク
- 20Webサーバすべてが，同一コンテンツのサイトを公開

これら20Webサーバは，明らかにSpamサイトであることである．したがって，べき乗分布からの突出した値に対応するWebサーバに関する具体的な調査は，今後の課題としたい．

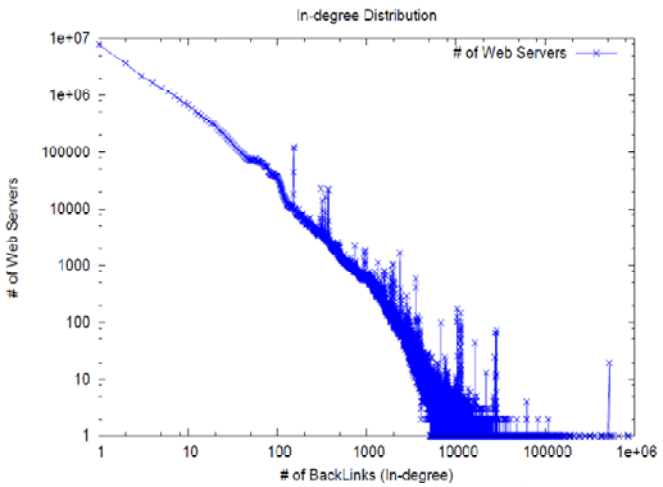

図 6　Web サーバの In-degree 値の分布

5. おわりに

本稿では，e-Societyプロジェクトで収集した約107億ページのページデータに対して，ccTLD毎のWebサーバの地理的設置位置の特定，Virtual Host数の地理的分布，およびバックリンク数（=In-degree数）の分布解析を行った．

その結果，国内設置Webサーバ数と国外設置Webサーバ数に相関関係があることを発見した．さらに，国外設置Webサーバ数が多いとして相関関係から外れたccTLDとして，以下の2つのような国々が検出された．

- 太平洋・インド洋・カリブ海に存在する小さな島国
- リビア・アフガニスタン・東ティモール等の国内問題を抱えている国

逆に，多数検出されたが，国外設置Webサーバ数が少ないとして相関関係から外れたccTLDはほとんど検出されなかった．これは，どのccTLDにおいても例外なく，一定以上の割合で，ドメイン管理国外にWebサーバが設置されていることを意味している．

また1ホストあたりに立てられているVirtual Host数が多いホストは，中南米，東欧諸国，バルカン半島に多く見られ，特にブラジル・ポーランドに設置されているホストが顕著であった．また，米国の内陸部，メキシコシティに設置されているホストもVirtual Host数が多いと検出された．

さらに，Webサーバのバックリンク数(In-degree値)の分布について計測を行ったところ，既存研究と同様に，Power-Lowの関係が成立していることを確認した．しかし，べき乗分布から外れた値が多数検出された．この外れ値はSpamサイトである確率が高いと考えられるため，今後は，べき乗分布から外れた値に対応するWebサーバを解析対象としたい．

[謝辞]

本研究の一部は，文部科学省リーディングプロジェクト「e-Society」 及び情報爆発プロジェクトとして実施した．

[文献]

[1] Hirate, Y., Kato. S. and Yamana. H.: “Web Structure in 2005”. In Proc. of WAW2006, LNCS 4936, pp. 36–46 (2008).
[2] Kamvar, D. S., Haveliwala, T. H., Manning, C. and Golub, G.: “Exploiting the block structure of the web for computing PageRank”, Technical Report, Stanford University (2003).
[3] Flake, G., Lawrence, S. and Giles, C.: “Efficient Identification of Web Communities”, In Proc. of 6th ACM SIGKDD Conf., pp.150-160 (2000).
[4] e-Society プロジェクト, http://www.yama.info.waseda.ac.jp/~yamana/es/
[5] 加藤真，山名早人: “Fact of the Web:30億ページのウェブの解析”, DEWS2006, 3B-i6 (2006).
[6] 平手勇宇，山名早人: “全世界のWebページのTLD・言語分布解析”，第70回IPSJ全国大会，5L-1 (2008).
[7] 童芳，平手勇宇，山名早人: “全世界のWebサイトの言語分布と日本語を含むWebサイトのリンク・地理的位置の解析”, DEWS2008, A2-3 (2008).
[8] 童芳，平手勇宇，山名早人:”全世界のWebサイトのTLD・言語分布・地理的設置位置の特定”，日本データベース学会論文誌, Vol.7, No.1, pp.31-36 (2008).
[9] Nakahira, K., Hoshino, T. And Mikami, Y.: “Geographic Location of Web Servers under African Domains”, In Proc. of WWW2006, pp. 989 － 990 (2006).
[10] 近藤浩之，手塚太郎，田中克己: “リンク元ページのアドレス情報に基づくWeb ページの地域的支持度の分析”, DEWS2007, B9-4 (2007).
[11] IP2Location, IP-Country-Region-City-Latitude-Longitude-ISP-DomainDatabase[DB8], http://www.ip2location.com/ip-country-region-city-latitude-longitude-isp-domain.aspx
[12] Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A. and Upfal, E.: “The Web as a Graph”, In Proc. of PODS2000, pp. 15－17 (2000).
[13] Barabasi, A. and Albert, R.: “Emergence of Scaling in Random Networks,” Science, Vol. 286, pp. 509－512, (1999).
[14] Broder, A., Kumar, R., maghoul, F., Raghavan, P. Rajagopalan, S., Stata, R. Tomkins, A. and Wiener, J.: “Graph Structure in the Web”, In Proc. of WWW2000, pp.309－320 (2000).
[15] Boldi, P., Codenotti, B., Santini, M. and Vigna, S.: “Structural Properties of the African Web”, In Proc. of WWW2002, Poster Session (2002).
[16] Liu, G., Yu, Y., Han, J. and Xue G.: “Gina Web Graph Measurements and Evolution”, In Proc. of APWeb2005, LNCS 3399, pp. 668－679 (2005).
[17] Netcraft Ltd, http://news.netcraft.com/

平手 勇宇　Yu HIRATE
2005 早大・理工学研究科修士課程修了．2008 早大・理工学研究科博士課程修了．博士（工学）． 2006より同大メディアネットワークセンター助手．ACM，IEEE，IPSJ，DBSJ各会員

片瀬 弘晶　Hiroaki KATASE
2008 早大・理工学部卒業． 現在，同大学大学院基幹理工学研究科修士課程在学中．Webリンク解析に関する研究に従事．IPSJ，DBSJ会員．

山名 早人　Hayato YAMANA
1993 早大・理工学研究科博士課程修了．博士（工学）．1993 - 2000 電総研．2000 早大・理工学部助教授．2005 同大理工学術院教授，現在に至る． ACM，IEEE，IEICE，IPSJ，DBSJ各会員．