

Web リソースの経年変化を用いた情報基盤利用状況の分析

難波 弘行

中平 勝子

三上 喜貴

長岡技術科学大学

1 はじめに

情報化の進展とともに Web 利用は世界的規模で拡大しているが、地域によってその利用状況は様々であり、局所的な発展や衰退には大きなアンバランスが見られる。この要因として、情報機器を購入するための経済力の格差、サーバ運営のための電力や通信回線の利用可能性といった様々な条件が地域毎に異なることがあげられる。

インターネットをはじめとする情報技術は地域間の格差を解消する切り札として期待されているにもかかわらず、実態としては、逆に持てる国と持てない国の格差を広げているおそれがある。その実態を、客観的かつ定量的に把握することは重要であり、筆者らは、こうしたデジタルデバイドを観測、評価する手法の研究を行っている。

本稿では、こうした指標化のひとつの領域として、ccTLD を地域を示す単位とし、各地域における Web の利用実態の指標化を試みる。ccTLD は、世界的に公平に割り当てられたインターネット・リソースのひとつである。しかし、実態としては、ccTLD が必ずしも ccTLD の示す地域で使われておらず、島嶼国を中心とした地域の ccTLD などほぼ国外の利用者によって利用されていることが指摘されている[1]。これは、該当地域における Web 資源の利用が該当地域におけるインターネット利用の真の姿を表していないばかりか、該当地域におけるインターネット利用の発展を逆に阻害する恐れもあり、該当地域の管理者による適切な管理が行われない場合には、地域外の利用者による違法な Web サイトやスパムメールの温床になる危険性もある。

こうした問題意識を踏まえ、ccTLD の利用状況の変化を継続的、客観的、定量的に観測する手法を開発することは、デジタルデバイド問題をより正確に分析する手段を提供すると考える。

2 方法

本研究では、継続的、客観的、かつ定量的に Web 利用の実態を観測するための一時的な情報源として、ロボットによるクロウリングによりデータ取得の可能な Web ページデータを基礎に分析を行った。

データの取得にあたっては、ミラノ大学の開発したクローラロボットである UbiCrawler を使い、2005 年と 2009 年に収集したアフリカ地域 57 カ国のデータを使用した。クロウリング・プロセス自体も多くの研究課題を含むため、2005 年と 2009 年では異なるパラメータ設定で行わざるを得なかった。特に 1 ホストから取得する最大ページ数については、通信事情の悪いアフリカ地域においては、観測対象サーバへの負担軽減の観点から、パラメ

ータを大きく変えざるを得なかった（2005 年は 1,000,000 ページ、2009 年は 5000 ページ）。このため、両年度間の取得データを比較するに当たっては、適切な補正を行う必要があり、特に後述するホストあたりページ数については、2009 年度のパラメータの条件にそろえるためにデータの補正を行っている。両年度ともホストあたりのページ数には同一の上限（5000）が存在し、その限度内での比較とならざるを得ないが、多くのサーバの総ページ数はこの上限内で分布していることを確認しており、比較の有効性はあるものとする。なお、これ以外のパラメータで相違するものは、データの取得速度にかかわるものであるが、この要因は年間比較の結果にはおおきな影響を与えないと考える。対象とする TLD は、アフリカ地域に属する 57 個の ccTLD に限定し、国の特定が困難である gTLD は分析対象外とした。また、サーバ所在地情報の取得には Maxmind 社が提供している IP アドレスから地理情報を得る仕組みである GeoIP を使い、各 ccTLD ごとに得られた各種データをそれらのサーバ所在地、特に当該地域の地域内か、地域外かによって分割した。

以上の手順によって得られた両年度のサーバ数（一意なホスト名の数とする）である S_{2005} , S_{2009} , ページ数 P_{2005} , P_{2009} と及びその国内・国外設置内訳の各データを用いて以下の分析を行った。

3 結果・考察

3.1 サーバ設置の国内外比率

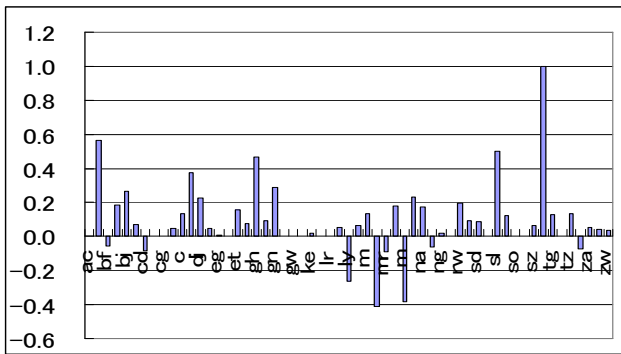
第一の指標は、該当地域における Web サーバの国外設置比率である。ある地域のサーバ管理者がサーバを国外に設置するに至る理由には様々なものが考えられる。その中には、国内における各種の規制（例えば当局による検閲や閲覧の制限など）を回避する目的なども含まれるものと想像されるが、主たる理由は、通信回線の事情、電力事情、サーバ管理のための技術スタッフの利用可能性といった、広義の通信基盤の不十分さを回避することであると思われる（この点についての検証は、3.3 項参照）。

図 1 は 2005 年から 2009 年における各 ccTLD ごとの Web サーバ国外設置比率の変化（2009 年の国外設置比率 - 2005 年の国外設置比率）をプロットしたものである。アフリカ地域全体として見ると、Web サーバの国外設置比率が増加傾向にあることがわかる。両年度間で、国外設置比率が低下したのは、調査対象である合計 57 の ccTLD のうち、わずか 8 つの ccTLD のみであった。アフリカ地域は、アジア地域に比べサーバの国外設置率が高いことがわかっているが、その傾向がより強まっているという事実は、アフリカ地域における情報基盤の整備の遅れが一向に改善されていないものであることを意味していると解釈できる。

An analysis of information infrastructure based on Web resources

Hiroyuki Namba, Katsuko T. Nakahira, Yoshiki Mikami
Nagaoka University of Technology

図1 サーバ国外設置割合の変化



3.3 サーバ国外設置比率が高い国の通信事情

上記の解釈の妥当性を検証するために、国内設置サーバ上のウェブページが観測されなかった地域について、その海底ケーブル接続状況を調査した。調査対象地域 57ccTLD のうち、2005 年、2009 年共に国内設置サーバからの Web ページが全く取得できなかった 10 カ国を表 1 に示す。これらのうち、島嶼が半数を占めている。残りの 5 ケ国は中央アフリカを除き海岸に面する国家であるがやはり海底ケーブル接続は行われていない。今日、国際インターネット接続は主として海底ケーブルで担われていることから、これらの地域では隣国などからの VSAT などを含む接続、あるいは衛星回線による帯域の限られた接続しか得られない。

表 1 国内設置サーバ上にウェブページが観測されない地域の海底ケーブル接続状態

分類	Cc	地域名	海底ケーブル接続の有無
島嶼	Ac	アセンション島 (英)	なし
	Io	インド洋地域 (英)	なし
	Re	レユニオン (仏)	なし
	Sh	セントヘレナ	なし
	St	サントメプリンシペ	なし
その他	Cd	コンゴ共和国	なし
	Cf	中央アフリカ	なし
	Bn	ギニア	なし
	Lr	リベリア	なし
	So	ソマリア	なし

3.2 1サーバあたりのページ数

第二の指標として、1サーバあたりのページ数について分析を行った。この指標は Web 利用の発展の度合いを示す指標と解釈することができる。特に、大規模なポータルやハブの成長、あるいはオーソリティーといわれる情報源サイトの成長によりこの指標は増大する。

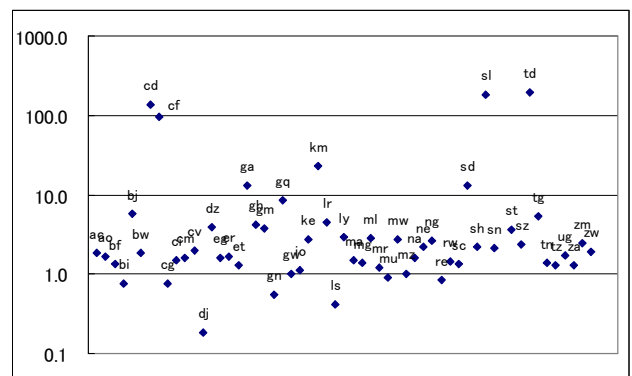
図 2 は、各 ccTLD の 2009 年の 1 ホストあたりのページ数を、2005 年の同様の数字に対する比率として表したものである。全体として、ほとんどの ccTLD でこの比率は上昇しており、概ね 20% の ccTLD で 5 倍を越える上昇が見られた。先述の指標とも関連するが、Web 利用者が海外である場合にも、その利用の高度化が進めばこの指標は上昇するために、この指標の上昇を単純に当該地域にお

ける Web 利用の発展と結論づけることは出来ないが、ひとつの指標ではある。このうち、特に顕著な変化 (100 倍以上) が見られた .cf, .td, .cd, .sl の 4 つの ccTLD について変化をもたらした要因の分析を行った。

- cf(中央アフリカ共和国), td(チャド)
2009 年には Google が参入していることが確認でき、google ドメインが該当 ccTLD のページのほとんどを占めている。
- cd(コンゴ民主共和国)
2009 年には香港にサーバを持つ巨大なポータルサイト (kats.cd) の出現が認められた。
- sl(シエラレオネ)

Google のほか 5000 ページを超える Web サイトが複数確認できた。しかし、いずれも国外に Web サーバを置くものであった。

図 2 1ホストあたりの Web ページ数の比較



このように、上記 4 つの ccTLD において見られた変化は、全て国外にサーバを置く Web サイトによるものであった。Google に代表されるようなポータルサイトの参入は、当該国・地域のユーザにとってみると、一面では好ましい変化ともいえるが、国内における開発活動の不足を物語るものでもある。

4 まとめ

本稿では、ウェブクロウリングによって収集したアフリカ地域の Web ページデータの経年変化を用い、デジタルデバイドの実態を観測するための手法確立のため、サーバの国外設置比率、ホストあたりのページ数という二つの指標について、その指標の経年変化状況の観察、変化の意味するものについての解釈や背景にある要因についての検証をおこなった。今回、使用した 2 時点のデータは、収集の条件が必ずしも同一でなく、事後的な補正を加えたものであること、分析した指標もごく限られたものであること等から、見出しえた結論も限られたものではあるが、今後、あらためてデータの取得、解析を行い、過去データとのより詳細な比較・分析を進めていく予定である。

参考文献

[1] Nakahira, K., Hoshino, T., And Mikami, Y., "Geographic Location of Web Servers under African Domains", In Proc. of WWW2006, pp.989-990(2006)