

## 各種メディアへのリンク解析に基づいた TLD オープン性評価

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

## 1.1 背景

インターネットのオープン性は、米連邦通信委員会[1]によって、ユーザが差別やコントロールを受けずに自らの意思・選択により自由にアクセスし、発言や創造活動を行うことができることであると定義されており、それが革新や経済成長にも繋がることも提唱している。また、Google や YouTube といった大手インターネット企業が加盟する Open Internet Coalition[2]は、オープン性はインターネットの基本原則であるという考えに基づき、オープンなインターネット実現のための取り組みを行っている。

近年、情報ネットワークの発達が目覚ましく、今日インターネットは我々の生活に欠かせないものとなっているが、インターネットによる恩恵を受けられる国と受けられない国の間での「デジタル・デバイド」が問題になるなど、世界的に見ると依然として格差が存在する。その格差は、主に通信ケーブルや機器の購入等、情報基盤の整備に依存する部分が大きく、これまで情報基盤格差を観測するための研究が行われてきた[3]。

しかし、情報基盤整備が徐々に進み基盤の問題が解消され、各国で Web 利用者が増えつつある一方で、検閲やフィルタリングなど、Web 利用を制限する新たな要素も増加しつつある。具体的には、アクセスブロック、検閲、フィルタリング等である。これらは、違法なコンテンツ防止策として Web 空間の規律を守る役割を持つが、一方で Web 利用の性質を変える網となってしまうこともある。近年のアラブ諸国や中国等で発生した民主化運動に伴うデモ活動と Facebook や Twitter といった各種メディアへのアクセス制限などのインターネット規制といった事例は、Web サービスが社会的影響力を持ち始めるとともに国家によるインターネットコントロールの対象になり得ることを示唆している。

こうした事象から、今後の健全な Web 利用を実現していくうえで、インターネットにおけるオープン性を国単位で評価することが重要であると考えられる。

本稿では Facebook や Twitter といったソーシャルメディアと呼ばれるメディア群に着目し、その利用度合いの推計に基づいた国別の Web 利用オープン性の評価を試みる。また、本稿では国を示す単位としては、ccTLD を用いるため、この評価結果を TLD オープン性と定義する。

## 2 方法

## 2.1 使用データ

本稿では、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

よるクロー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ーロボットである UbiCrawler[4]を用い、2009 年に収集したアジア地域に割り当てられている 49 の ccTLD を持つ Web ページを収集した。

ソーシャルメディアの利用を測るために、本稿ではハイパーリンクを用いる。インターネット上の Web ページに記述されるハイパーリンクは、特定のページへ移動するために記述されるものであり、いわば道路のような物である。このため、実際の利用の度合いがリンク数に反映されると考えられる。

そこで、収集した Web ページに張られているリンクから各ソーシャルメディアに対するリンクのみを抽出し、ccTLD 毎に各ソーシャルメディアへのリンク数を集計した。なお、本稿では調査対象ソーシャルメディアとして世界で広く使われており、多様な言語に対応している YouTube, Facebook, Twitter, Wikipedia の 4 つのソーシャルメディアを選定した。また、各ソーシャルメディアの主な機能と利用可能な言語数を表 1 に示す。

表 1 各ソーシャルメディアの機能と対応言語

サービス名	主な機能	対応言語数
YouTube	動画投稿・閲覧	54
Facebook	Web 上コミュニケーション コミュニティ形成	77
Twitter	短文による発言 他者発言の閲覧	22
Wikipedia	情報・知識の参照・公開	273

## 2.2 オープン性指標の定義

Web 上でのリンク活動を行う主体を人間としたとき、ある特定の Web ページへのリンク数  $L$  を、観測可能な次の変数である人口  $N$ 、インターネット利用者数  $P$ 、インターネット利用水準  $U (=N/P)$  を用いて表すことを考えたとき、 $L$  は式 (1) の様に表すことができる。

$$\begin{aligned} L &= N \times \frac{P}{N} \times \frac{L}{P} \\ &= N \times U \times \frac{L}{P} \quad \dots (1) \end{aligned}$$

ここで  $L/P$  はインターネット利用者のリンク活動の自由度と捉えることができる。そこで、オープン性指標  $O$  を

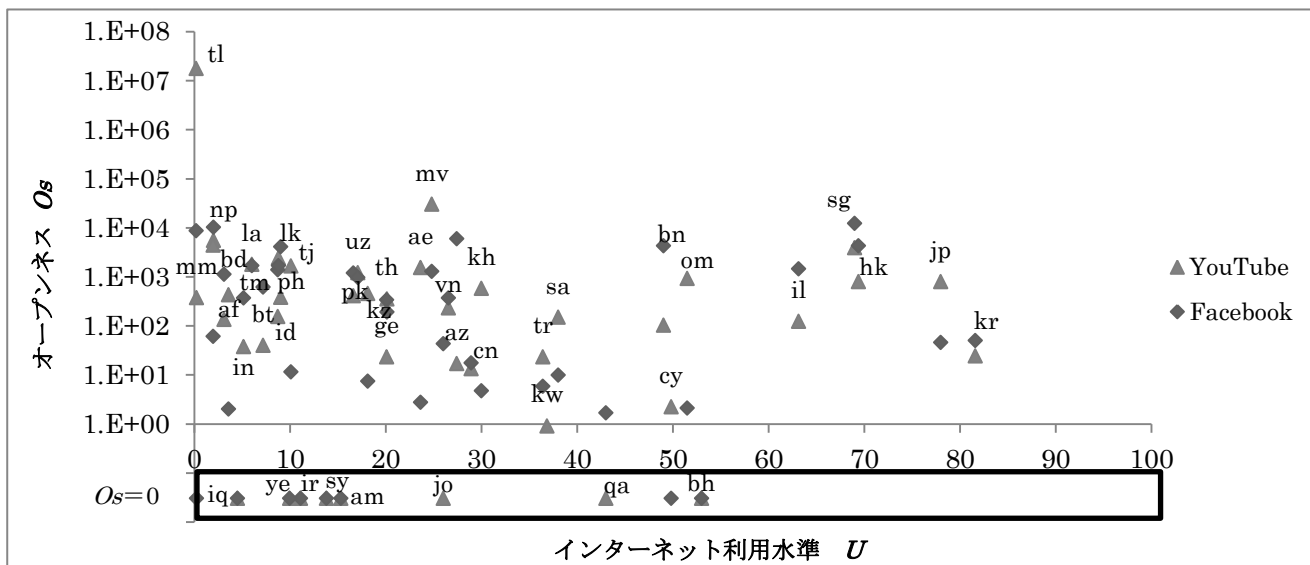
$$O = \frac{L}{P} \quad \dots (2)$$

と定義することができ、本稿では (2) 式をオープン性指標とする。上記の方法で  $O$  を算出した上で、 $O$  と  $U$  の関係をグラフに示し、考察を行う。

The estimation of TLD openness based on analysis of link to various media web sites

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

図1 インターネット利用水準  $U$  とオープン性  $O_s$  の比較



また、以降の解析では、特定ソーシャルメディアへのリンクやオープン性については添え字  $s$  を各変数につけて表すこととする。

### 3 結果・考察

2の手順で求めた各 ccTLD の  $O_s$  のうち、例として Facebook と Youtube の  $O_s$  について各国の 2009 年におけるインターネット利用水準  $U$  と比較したグラフを図 2 に示す。

図 2 の全体を見ると、主に  $U$  が 60 以下の ccTLD において  $O_s$  が極端に高いものと低いものが混在しており、 $U$  が 60 を超える ccTLD では  $O_s$  が比較的高い値で安定している傾向が見られる。以降で、この結果に対する考察を行う。

#### (1) $U$ が 60 以下の ccTLD

$U$  が低いにも関わらず高い  $O_s$  を示す ccTLD が多く出現した。これは、ccTLD のユーザが必ずしもその ccTLD の帰属国の者とは限らないということに起因していると考えられる。例えば、 $U$  が低い国の ccTLD でも、 $U$  が高くソーシャルメディアをよく利用する国外ユーザがその ccTLD を活発に利用しているとすれば、そのギャップによって  $O_s$  の値が釣り上げられてしまうことは大いにあり得る。実際、Youtube の  $O_s$  で以上に高い値を示した tl (東ティモール) について元データを調査したところ、ある特定の Web サイトからのみ大量に Youtube に対するリンクが張られており、なおかつその Web サイトは東ティモールではほぼ話されていない言語で記述されたものであった。このため、このようなケースについては、リンク元の Web サイトが記述されている言語と、当該国で利用されている言語と相違を見るなどして、純粋な当該国ユーザによる利用であるかの検証を行う必要がある。

一方、ソーシャルメディアへのリンクが全く確認できなかった ccTLD として、ir (イラン) や sy (シリア) のように、国境なき記者団[5]によって「インターネットの敵」とされている国の ccTLD が確認でき、このような国については規制やコントロールによってソーシャルメディアを自由に

利用できるような環境が存在しておらず、それが今回の結果に表れたという可能性が考えられる。

#### (2) $U$ が 60 を超える ccTLD

$U$  が 50 を超える国の ccTLD の  $O_s$  については、ソーシャルメディアの種別による極端な  $O_s$  のばらつきも見られず、比較的高い値で安定しているため、こういった国の場合はオープン性が担保されていると考えられる。

このため、 $U$  が 60 を超えるような国の ccTLD において低い  $O_s$  が検出された際には、国家によるコントロールの存在や異常な事態を疑うべきであると考えられる。

また、今回の結果において、 $O_s$  の値が  $10^4$  を超えることは一部を除きほぼ無かったため、 $O_s$  の上限値はおおよそ  $10^4$  前後であることが予想でき、基本的に  $O_s$  の値はその上限値に向かい、安定していくものであると考えられる。

### 4 まとめ

本稿では、ウェブクロールによって収集したアジア地域の Web ページ上の各種ソーシャルメディアに対するリンクを用いて、ccTLD 毎のオープン性評価を試みた。

今回の結果からは、 $O_s$  の値がソーシャルメディアの利用を反映する様子が確認できた一方で、インターネット利用水準の低い国を中心に実際の利用を示しているかといった不明確な部分も残ってしまった。このため、今後は利用言語のような Web ユーザの利用を左右する要因も考慮した上でオープン性の評価を進めていく予定である。

#### 参考文献

- [1] Federal Communications Commission, The Open Internet.
- [2] Open Internet Coalition <http://www.openinternetcoalition.org/>
- [3] Nakahira.K, Hoshino.T, And Mikami.Y, "Geographic Location of Web Servers under African Domains", In Proc. of WWW2006, pp.989-990(2006)
- [4] Ubcrawler <http://law.dsi.unimi.it/>
- [5] Reporters Without Borders <http://en.rsf.org/>