

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 23 日現在

機関番号：13102

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500308

研究課題名(和文) Webリンク解析によるインターネット空間のオープン性評価手法

研究課題名(英文) proposal of assessment Openness in the Internet derived from Web link analyzation

研究代表者

中平 勝子 (Nakahira T., Katsuko)

長岡技術科学大学・工学(系)研究科(研究院)・助教

研究者番号：80339621

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：インターネットサービスへ容易にアクセスできるか否かは情報格差や知識格差につながり得る。その実態を調査するには、インターネットの中立性(オープン性)やドメイン管理状態を可能な限り現実のインターネット上で観測し、その評価指標を構築する必要がある。本申請では、実際のユーザの立場でインターネットのオープン性評価を行うため、自由なアクセスの代表格と考えられるソーシャルメディアへのWebページからのリンクと、リンク元が記載される言語の関係について観測し、その結果から各地域のユーザがどの程度ソーシャルメディアへのアクセスが容易であるかを判定する指標を三角図として提案した。

研究成果の概要(英文)：The ease of access to the Internet is connected to the digital/knowledge divide so that there is rich and free information in it. To observe and assess these phenomena in the Internet, we propose a new assessment index for Openness in the Internet and its management. We assumed that the fact "a Web page has links to social media" should be the evidence of "openness" of the Internet of the region where the Web page is located. Since the Web page is written in some language, we could estimate the degree of openness of the region in terms of presence/absence of the links to social media of the Web pages in the region. Figuring ternary diagrams consists of three types of language (official, local, foreign) and the openness value shows the assessment for ease accessing social media.

研究分野：教育工学/情報学基礎

キーワード：インターネット オープン性 リンク解析

1. 研究開始当初の背景

インターネットにおける新たなサービスが拡大する中、新サービスの享受可否が新たな情報格差や知識格差を生み出す可能性がある。その実態、すなわち、インターネットの中立性/オープン性を調査するには、インターネット社会の実測および評価指標を構築する必要がある。インフラに対する整備ポリシーについてはアメリカにおけるインターネットへの合法的な接続・コンテンツ送受信をはじめとするユーザ側の権利保証といったものが提唱されているが、あくまで提唱であり、評価指標の構築が難しい実態がある。

2. 研究の目的

SNS やニュースメディアなど特徴あるサイトへの Web リンク解析を行うことにより、インターネット空間のオープン性を評価する手法を開発する。Web リンクは合法的かつ自由なインターネット利用の維持のための重要な情報として、インターネットにおけるユーザ活動の自由度に即したオープン性を示す客観データと考えることができ、対人口リンク数などからオープン性評価について検討する。また、発展的に、これらを担保する客観データに基づいたドメイン管理の中立性、レジリエンス(回復性)などについても推定を行う。

3. 研究の方法

Web サイト本体から推定できるインターネットのオープン性と、全体的なドメイン管理およびそのレジリエンスの推定、およびそれらの事象を捉えるための枠組みの記述、という3つのアプローチから研究を行った。

(1) インターネットのオープン性

Web サイトからの特徴あるサイトへの Web リンク解析を行うにあたり、もっとも重要なものは Web ページ収集のためのクロール環境の整備である。これには、言語天文台のリソースをそのまま用いたが、老朽化した機材についてはリプレースを行い、おおよそ 10TB 程度のデータをローカルに保持した上で、カセット HDD に過去のデータをストックすることで過去のクロールデータについては再追跡可能な環境を整備した。また、得られたデータについては、将来的に公開を予定しているため、そのデータシェア方式、特に解析用データ生成、分析ツールの構築についても検討した。以上の環境整備の元、言語の多様性が認められる地域、およびそれが認められない地域(アジア、ラテンアメリカ、ロシア)について、実際に Web サイトに記載される SM へのリンク、ここでは Facebook への SM リンクについてリンク数 L_i 、インターネット利用者数 N_i 、インターネット普及率 U_i を用いて、ある国 i におけるオープン性 O_i を $L_i/(N_i U_i)$ で算出することとした。そこに、リンク元ページで使用されている言語について公用語/local 言語/その他の外国語、に区別し、

それらとの関係を三角図として示し、その特徴を見出す手法をとった。

(2) ドメイン管理およびレジリエンス推定

全体的なドメイン管理およびそのレジリエンスの推定には、Web リンクだけでは実態を把握しきれないところもあるため、発想を変え、スパムメール送信サーバの分析という観点から推定を行う手法を開発することとした。これは、スパムメール送信サーバが送信するスパムメール送信パターンを分析することで、該当サーバの管理状態を推定し、それを ccTLD 毎にまとめ、サーバの行動が不審になった状態をどの程度の時間経過で察知し、元の正常利用な状態に戻せるか、の推定を3つの指標を用いて定性的に推定することとした。

(3) これらの事象を捉えるための枠組み

以上の観測結果および推定結果における、一定の整合性は、これらを含めたネットワーク社会を合理的に記述できる枠組みを用いて論証するのがこの場合にはふさわしい。そこで、e ネットワークという枠組みを導入し、提案したオープン性やドメイン管理の観測手法について一定の理由付けが可能であるかの検証を行った。

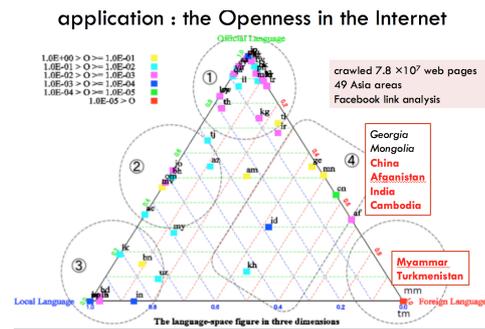


図 1 インターネットのオープン性と使用言語の関係

4. 研究成果

(1) インターネットのオープン性

典型的な例を図 1 に示す。アジア諸国における Web ページからの SM リンクに基づくオープン性指標と使用言語の関係上部頂点が公用語、左下頂点が local 言語、左下言語がその国ではおおよそ使われていないであろう外国語である。図中、領域 1 から領域 3 は、local 言語や公用語を含むことから、ほぼその国内で使用されている、すなわち、Web ページを持つ国の民自身がインターネットの利益を享受できているとみなす。領域 4 については国外で使われる言語の Web ページからのリンクが多いことから、Web ページを持つ国の民自身がインターネットの利益を享受しづらい状況にあるとみなす。領域 5 については、国外言語で書かれたページからのリンクが多いことから、ほぼその国では Facebook 等 SM への接続ができない、すなわちオープ

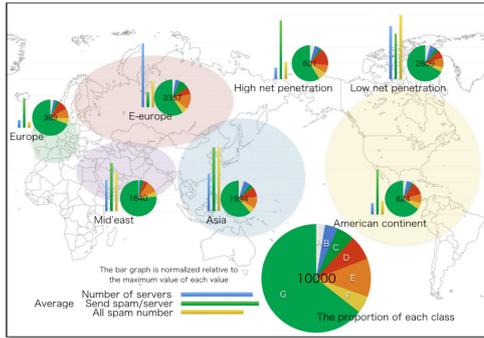


図2 地域ごとのスパム送信特性

性が担保されない国であるという実態を示すことができた。研究推進中、これらの国についても SM へのアクセス拒否を部分的に解除するニュースが流れた。

(2) ドメイン管理およびレジリエンス

ドメイン管理そのものの本来の様子は、ドメイン管理者に状況を聞くのがもっとも確実な方法である。しかし、ネットワーク管理についてはオープンにできる部分とそうでない部分があることから、各ドメイン配下に存在するサーバ群自身から短時間に push 形式で発信される情報として捉えられるメールの送信状況に着目した。メール送信には、正常なメール送信とスパムメール送信の2種があるとし、この内スパムメール送信行為は本来、存在しないほうが良い不審行為であると考え、そうした行為を即座に停止できない度合いで以って、近似的に個のサーバ管理状態、すなわち、常時モニタされている(そうした技術が導入されている)か否かを推定する。その推定には Evolution Diagram という手法を構築し、最大送信能力、継続性、再起性という3つの指標でサーバのレジリエンスを評価する。これら個のサーバは、いづれかの国/地域に属しているため、地域ごとのサーバ管理状態を推定することで、健全なドメイン管理がなされている、もしくはレジリエンスが強い状態であるかそうでないかの判定を下し、7つのレジリエンスクラスに分類した。その成果の一部を図2に示す。全体的に、緑(ほぼセキュア)なサーバ群であるが、そうではない部分については地域によって特徴が見えることがうかがえる。ただし、観測地点が1箇所しかないため、今後こうしたレジリエンス測定を各地域で行うことでより精緻な実態へとつながると考える。

(3) 事象を捉えるための枠組み構築

以上、2種類の観測結果を用い、インターネットのオープン性を捉えるための手法を構築できたと考えているが、なぜその手法が良いのか、については理論的な研究が必要である。そこで、図3に示すeネットワークを

用いてインターネット上に起きている諸現象を説明することを試みた。この枠組みについては研究期間終了後も継続してその合理性を検証する必要がある。eネットワークは、ネットワーク社会を説明するために必要な4要素(人的要素、基盤要素、環境要素、生成物)およびそれらを繋ぐメディアで社会現象をシンプルに説明するための枠組みである。これを用いることで、オープン性は人的要素-基盤要素-メディアで構成される面現象に社会制度を含めた環境要素が関与するものとして、ドメイン管理とレジリエンスについては、人的要素-基盤要素-環境要素の結果生じる現象が生成物であるスパムメールを通して観察されるもの、として規定できることを示した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4件)

上村圭介, "国別トップレベルドメイン名の利用促進要因の推定と統治体制の特徴抽出", 情報社会学会誌 Vol. 7, No. 2, pp23-40, 2013 (査読有)。

<http://infosocio.org/vol7no2-02.pdf>

Tin Htay Hlaing and Yoshiki Mikami, "FSA Based Code Sequence Checking to Prevent Mal Use of Myanmar IDNs", International Journal of Computer Science Issues, Vol. 11, pp. 25-34, 2014(査読有)。

[Http://www.ijcsi.org/papers/IJCSI-11-1-1-25-34.pdf](http://www.ijcsi.org/papers/IJCSI-11-1-1-25-34.pdf)

Katsuko T. Nakahira, Kakeru Yamaguchi, and Muneo Kitajima, "Ecology of Spam Server Under Resilience Force in the e-Network Framework", The seventh International Conference on Advanced Cognitive Technologies and Applications, 169-175, 2015 (査読有)。

Katsuko T. Nakahira, "Understanding social phenol in cyberspace: focusing on language, infrastructure, and contents", 3rd International Conference on Linguistic and Cultural Diversity in Cyberspace, to be publicshed, 2014 (査読無)。

[学会発表] (計 18件)

Katsuko T. Nakahira, "A Framework for Understanding Human eNetwork - Interaction among Language, Governance, and more", III Symposium international sur le multilinguisme dans le cyberspace, 2012. 11. 21-23, Paris.

山口翔生, 中平勝子, 北島宗雄, "スパムメ

ールサーバの発信行動変容の推定”，情報科学技術フォーラム，2012.9.4，法政大学（小金井市）。

中平勝子，北島宗雄，”人の営みとして形成されるeネットワークのダイナミクスを解明するための枠組み”，ARG WE2 研究会，2012.12.14，神奈川近代文学館（横浜市）。

Omer Osman Ibrahim and Yoshiki Mikami，“Indexing the Tigrinya Web”，情報処理学会全国大会，2013.3.7，東北大学（仙台市）。

細井亮司，中平勝子，北島宗雄，”情報取得の過程を考慮したWebサイト上におけるオブジェクト選択の観察”，情報科学技術フォーラム，2013.9.5，鳥取大学（鳥取市）。

山口翔生，中平勝子，北島宗雄，”メール送信サーバ情報送信量特性”，情報科学技術フォーラム，2013.9.6，鳥取大学（鳥取市）。

Katsuko T. Nakahira，“Web-Based Vulnerable Peoples - Focusing on Language--”，Internet and Socio-Cultural Transformations in Information Society，2013.9.8 - 12，Yuzno-Sakhalinsk，Russia。

伊藤公，中平勝子，三上喜貴，”デジタルデバイス研究/分析用DBの構築”，電子情報通信学会信越支部大会，2013.10.5，長岡技術科学大学（長岡市）。

佐々木拓美，中平勝子，”情報流通に着目したデジタルデバイス構造”，電子情報通信学会信越支部大会，2013.10.5，長岡技術科学大学（長岡市）。

伊藤公，中平勝子，三上喜貴，”国別ドメイン利活用分析のためのプロビジョンスキーム”，情報処理学会全国大会，2014.3.12，東京電機大学（東京都特別区）。

佐々木拓美，中平勝子，“基盤・ソーシャルメディアに着目したインターネット構造の変化”，情報処理学会全国大会，2014.3.12，東京電機大学（東京都特別区）。

山口翔生，中平勝子，北島宗雄，”カントリードメインガバナンス推定を目指したスパム送信サーバの地域別特性分析”，電子情報通信学会総合大会，2014.3.19，新潟大学（新潟市）。

伊藤公，中平勝子，三上喜貴，“クローラデータのプロビジョンスキームにおけるファイル入出力機構の検証”，情報科学技術フォーラム，2014.9.3，筑波大学（つくば市）。

岩田翔汰，中平勝子，北島宗雄，“通信基盤

と人の相互作用による情報生成・流通プロセスの記述とバイラルメディアへの適用”，情報科学技術フォーラム，2014.9.4，筑波大学（つくば市）。

細井亮司，中平勝子，北島宗雄，“「意図」に着目したソーシャルメディア上の人-コンテンツ相互作用の記述”，情報科学技術フォーラム，2014.9.5，筑波大学（つくば市）。

山口翔生，中平勝子，北島宗雄，“スパム送信サーバの重篤度によるスパム送信パターン分類”，情報科学技術フォーラム，2014.9.5，筑波大学（つくば市）。

伊藤公，中平勝子，三上喜貴，“大容量データ流通のためのファイルシステムの開発・評価”，情報処理学会全国大会，2015.3.19，京都大学（京都市）。

山口翔生，中平勝子，北島宗雄，“行動変容過程の解明のためのスパム送信サーバ観測”，情報処理学会全国大会，2015.3.18，京都大学（京都市）。

〔図書〕（計 1 件）

三上喜貴，中平勝子，児玉茂昭，慶應義塾大学出版界，“言語天文台からみた世界の情報格差”，336p，2014。

〔産業財産権〕

○出願状況（計 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
取得年月日：
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

中平 勝子 (Katsuko T. Nakahira)
長岡技術科学大学・工学部・助教
研究者番号：80339621

(2) 研究分担者

三上 喜貴 (Yoshiki Mikami)
長岡技術科学大学・工学部・教授
研究者番号： 70293264

(3) 研究分担者

北島 宗雄 (Muneo Kitajima)
長岡技術科学大学・工学部・教授
研究者番号： 00344440

(4) 研究分担者

上村 圭介 (Keisuke Kamimura)
大東文化大学・外国語学部・准教授
研究者番号： 10319014