

カントリー・ドメイン・ガバナンス分析用 統合データ管理システム

新井 裕樹 中平 勝子 三上 喜貴

長岡技術科学大学 工学研究科

1. はじめに

URL を構成する ccTLD は世界中の国・地域に平等に割り当てられた資源であり、現在約 250 の ccTLD が存在する。ccTLD は、本来 IANA によって割り当てられた国のユーザが使用することを想定されたものであるが、これを自国ユーザではなく、主に海外のユーザに提供するケースも少なくない。

ccTLD の管理に当たる ccTLD 管理者の配布政策、料金、規律や技術的能力、組織体制等には大きな差異が存在している。その差異は配布政策や料金の差異は当該国・地域におけるインターネット利用の水準に大きな影響を与え、また規律や技術的能力の欠如は当該ドメインがスパムや不適切なコンテンツの温床となっている。安全・安心かつ地球規模でユビキタスなネットワークの実現のためには、すべての ccTLD 管理者が一定の水準を満足するカントリー・ドメイン・ガバナンスを実現する必要がある。

筆者らはこうした指標を開発するに当たって言語天文台[1]、サーバの地理的所在地[2]、島嶼国 ccTLD 利用実態[3]、情報格差指標[4]、リンク構造解析[5]の 5 つの研究を行ってきた。これらの研究は、いずれもカントリー・ドメイン・ガバナンスの分析を目的としたものであり、また、そのために独自の分析用指標、ソフトウェア・ツールを開発している。しかし、これらの研究は一定の問題意識を共有しつつも、分析のために開発されたソフトウェア・ツールはシステムとして統合されていない上に、指標を継続的に計測できるような構成となっていない。このため、指標を継続的に常時分析できるシステムとして統合する必要がある。

本稿では、この統合システムで取り扱う指標“CDG Maturity Index”の構想、及び開発中のシステム概要について報告する。

2. CDG Maturity Index

カントリー・ドメイン・ガバナンス (CDG)

Integration data management system for country domain governance analyses

Yuki Arai, Katsuko T. Nakahira, Yoshiki Mikami
Nagaoka University of Technology

の実態を ccTLD 単位として多面的に評価するための指標である CDG Maturity Index を説明する。CDG Maturity Index は 7 つの評価項目と調査データと独自分析データの二つのデータ属性に分けられる。調査データは、公的機関のデータを収集したもので ISC のドメイン数[6]やドメイン販売料金・方針などの指標がある。独自分析データは、インターネット上で実際に流通しているデータである。本システムでは独自分析データだけを取り扱うため、そのデータを使用した指標を本稿で取り上げる。10 個の指標から構成される CDG Maturity Index を表 1 に示す。

表 1 CDG Maturity Index の構想

評価項目	独自分析データ
[Accessibility]	通信パフォーマンス
[Affordability]	ウェブサーバ数・ページ数 サーバ国内設置比率
[Openness]	海外報道機関へのリンク率 リンク数の冪乗則適合度
[Secure]	リダイレクションページ比率
[Trusted]	空ページ比率
[Open Contents]	ドメイン結合度
[Diversity]	母国語ページ比率 文字コードの混乱状況

3. システム概要

本システムの概要を図 1 に示す。大きく入力部・出力部、およびそのインタフェースから構成される。また、継続的運用を意識したため、特別に管理部を設けている。以下、本システムを構成する各部を概説する。

3. 1 インタフェース

インタフェースは Web サーバ上で動作するアプリケーションとして実装する。ユーザがデータの入力、出力など、システムにアクセスするために使用される。

3. 2 入力部

収集フェーズはインタフェースで入力された情報を元にインターネット上からデータ収集するためにクローラを使用するフェーズである。

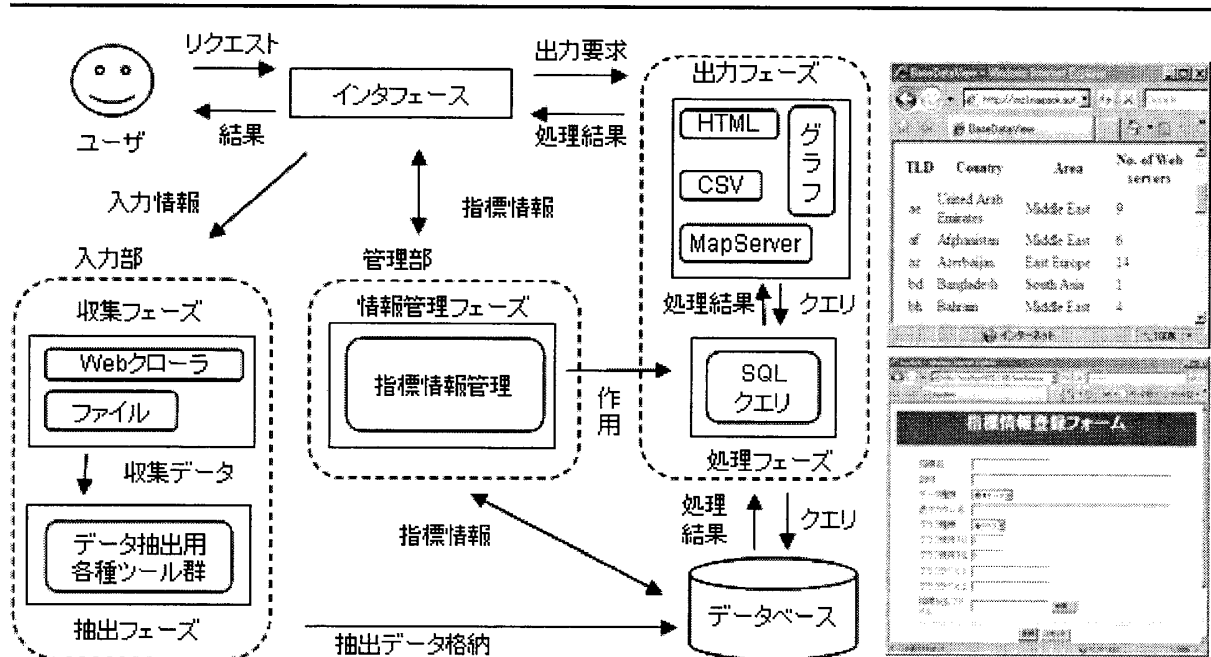


図1 システム概要図 (左) と出力画面例 (右上) と指標管理画面 (右下)

ページ情報を収集する UbiCrawler[7]とサーバ情報を収集するサーバ巡回クローラ[2]の二つによるデータ収集方法に加え、ファイルからデータを入力する方法がある。

抽出フェーズは、収集したデータから分析に必要なデータだけを抽出するフェーズである。抽出フェーズでは Ubi-API[7], HTML パーサ, LIM[1], GeoIP[8]のツールを用いて必要となるデータを抽出・解析する。

3.3 出力部

処理フェーズは、要求のあった SQL クエリを実行し、処理結果を返すフェーズである。指標を計測するための SQL クエリは、データベースに蓄積する構成となっている。そのため、指標の出力が要求されるたびにデータベースに格納されている SQL クエリを元に指標の計測を行う。

出力フェーズは指標を出力するための形式 (HTML, CSV, グラフ, MapServer[9])に加工し、ユーザに情報を与える。指標の出力例を図1の右上に示す。

3.4 管理部

管理部は情報管理フェーズの指標情報管理から構成されている。指標情報管理は、Web 上から指標情報 (指標名, 指標計測用の SQL など) を追加, 変更, 削除できる管理機能である。

本システムの提供する指標は、政策提言の参考情報や政策効果のフォローアップ指標として使用することが目的であるため、継続的にシステムを運用していくことが必要である。この過程で、指標の算出方法や指標体系の構成に変更が生じることもありうる。その際には原データを保持しておき、新しい指標算出方法に基づい

て再計算することにも対応する必要がある。本システムは指標の算出方法も、プログラムを変更することなくユーザが新しい SQL クエリの形で計算式を再定義するだけで対応できるようにした。指標管理画面例を図1の右下に示す。

4. まとめ

本論文では、カントリー・ドメイン・ガバナンス分析用統合データ管理システムについて述べた。今後は、カントリー・ドメイン・ガバナンスの実態を分析・評価していきたい。

参考文献

- [1] 児玉茂昭, チューチャーチョーン, 三上喜貴: 自動言語判定手法の開発とそれを利用したインターネット上の言語分布に関する調査, 日本語学会第135回大会, 松本, 2007.
- [2] Katsuko T. Nakahira et. al, Low-Load Server Crawler: Design and Evaluation, The 17th International WorldWide Web Conference, Beijing, 2008.
- [3] 和田祥太, 中平勝子, 三上喜貴: 島嶼国 ccTLD の利用実態, 電子情報通信学会信越支部大会, pp152, 2006.
- [4] 上嶋智大, 中平勝子, 三上喜貴: デジタルデバインドの評価指標についての一提案, FIT2007, pp481-484, 2007.
- [5] 石原直幸, 中平勝子, 三上喜貴: ccTLD を単位とした Web コミュニティ構造の分析, FIT2007, pp115-118, 2007.
- [6] Internet System Consortium, <http://www.isc.org/>.
- [7] UbiCrawler, <http://law.dsi.unimi.it/>.
- [8] GeoIP, <http://www.maxmind.com/>.
- [9] MapServer, <http://mapserver.gis.umn.edu/>.