

IP 地理位置情報を付加したアクセスログの時間的空間的分析システムの提案

渡邊 裕一[▼] 服部 哲[▼] 速水 治夫[▼][▼]

[▼]神奈川工科大学情報学部 [▼]神奈川工科大学大学院工学研究科

あらまし Web サイト運営者にとって効率的運営の重要性は日々増しており、アクセスログを分析することは有効的である。アクセス分析ツールを用いることでアクセスログを集約表示し、閲覧者の情報を獲得することができるが、閲覧者の地理位置情報の獲得と時間的空間的集約操作による閲覧者の地域の直感的な把握が困難という課題がある。そこで本論文ではアクセスログの時間的空間的可視化を目的として地理位置情報データベースを用いてアクセスログを時間的および空間的に集約しインタラクティブに分析可能なシステムを構築した。また地図インターフェースを用いることで時間的推移における閲覧者の地域を直感的に把握することができた。

A web access log analyzing system using IP address geolocation database

YUICHI WATANABE[▼] AKIRA HATTORI[▼] HARUO HAYAMI[▼][▼]

[▼]Information Faculty, Kanagawa Institute of Technology

[▼]Graduate School of Engineering, Kanagawa Institute of Technology

Abstract Efficient management for the website is increasing day by day. The website manager has to analyze web access log because they want to know website watcher's information. A web access log analyzing system can consolidate and show web access log. As a result, website manager can get the website watcher's information. But it has two problems. The first is to get watcher's geolocation is difficult because the web access log has not geolocation. The other is difficult to use system by time or region. Then we propose a web access log analyzing system using IP address geolocation database. Both this system and map interface could visual watcher's region and access count of the region by time.

1. はじめに

近年、Web サイト利用者は増加しており、Web サイト運営者にとっても効率的運営の重要性は増してきている。そこで Web サイト運営者は Web サーバにあるアクセスログを分析することで情報発信の成果を得ようとしている。しかしアクセスログには閲覧者の都道府県名や緯度経度など地域の情報が含まれていないため、地理位置情報の獲得が困難という課題がある。また著者らは閲覧者の地域が把握できるアクセス分析ツールを試用した。しかしアクセスログの期間を選択して閲覧者の地域を詳しく直感的に把握できないなど、時間的空間的集約操作による閲覧者の地域の直感的な把握は困難という課題もあった。

そこで本論文では地理位置情報データベース^①と地図インターフェース^②を用いてアクセスログ

から時間的空間的条件を考慮して閲覧者の地域を地図上に可視化するシステムを提案する。

以降 2 章では関連研究を述べ、3 章では試作システムの概要を述べる。4 章では試作システムの IP 地理位置情報変換付加モジュールについて、5 章では時間的空間的分析モジュールについて述べる。6 章ではアクセスログの可視化に用いる地図インターフェースを述べ、7 章で評価実験を述べる。最後に 8 章のおわりでまとめと課題について述べる。

2. 関連研究

本章では関連研究およびアクセス分析ツールの種類や従来のアクセス分析ツールについて述べる。関連研究では閲覧者の情報を得るために様々な観点からアクセスログを分析しようとす

る動きがある。文献(1), (2)ではアクセスログからデータ宝石箱アルゴリズムを用いてサイトマップ上に閲覧者のアクセス数を可視化しており、Web ページにおける閲覧者数が把握できたと報告している。また文献(3)では芸術作品として地図インターフェースを用いており、閲覧者の情報を地図上に表現する点、地理位置情報データベースを用いる点が本研究のヒントとなった。またアクセスログを分析するためのツールもいくつか存在する。アクセス分析ツールは大別するとサーバログ型と Web ビーコン型がある。サーバログ型は Web サーバに保存されたアクセスログファイルを分析する方式である。この方式は、記録されている過去のログファイルから情報を獲得できる点が特徴である。Web ビーコン型はサーバログ型と情報の収集方法が異なる。この方式では閲覧者が Web ページにアクセスするたびに、情報を獲得する。リアルタイムで閲覧者の情報を獲得・分析できるのが特徴である。

サーバログ型アクセス分析ツールとして Circle氏が作成した「ApacheLogViewer」がある。このツールでは、NCSA 形式のアクセスログファイルを項目ごとに集約して表示を行うことができる。特徴としては特定の IP アドレスがどのページを閲覧したのかを簡単に集約することができる。しかし閲覧者の地理位置情報の可視化を目的としていないため閲覧者の地域を直感的に得ることが困難である。

一方、Web ビーコン型アクセス分析ツールとして Google 社が提供する「GoogleAnalytics」がある。このツールでは付属している時間インターフェースと地図インターフェースにより分析期間における閲覧者の地域を市町村レベルまで可視化することができる。しかし時間軸を操作しながら閲覧者の地域を表示することが難しいため、閲覧者の地域を直感的に把握することが難しい。

本論文では地理位置情報データベースを用いてアクセスログに閲覧者の地理位置情報を加え、時間的空間的に分析・可視化するサーバログ型のアクセス分析ツールを構築する。

3. 試作システム概要

本論文で提案する試作システムは NCSA 形式アクセスログを入力として地図インターフェースである GoogleEarth の読み込みファイルである KML(KeyholeMarkupLanguage)ファイルを出力する。出力された KML ファイルを地図イン

ターフェースに読み込ませることで閲覧者の地域を時間スライダ機能と併に地図上に可視化する。

試作システムは 2 つのモジュールで構成した。一つ目はアクセスログの IP 情報を地理位置情報に変換しデータベースに記録する IP 地理位置情報変換付加モジュールである。二つ目は記録されたデータベースから時間的空間的条件を考慮して集約された KML ファイルを作成する時間的空間的分析モジュールである。システムの概略図を図 1 に示す。

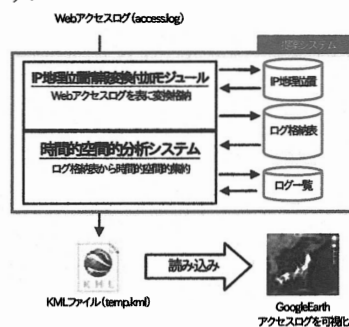


図 1 試作システム概略図

本システムの動作内容を述べる。ユーザはまず本システムのインターフェースを用いてアクセスログをサーバに送信する。その後、IP 地理位置情報変換付加モジュールを用いてアクセスログに地理位置情報を付加してデータベースのログ格納表に格納する。その後、ログ格納表から SQL を用いて、アクセスログを取り出す形で時間的空間的分析モジュールを用いる。時間的空間的分析モジュールでは時間的空間的に条件を選択して KML ファイルを作成できるようにした。作成された KML ファイルでは GoogleEarth の時間スライダ機能を有効にし、アクセスログの時間的空間的可視化を可能にした。本システムに適用した神奈川工科大学の公式 Web サイト⁶⁾のアクセスログの 2008 年 12 月 9 日の九州地方における閲覧者の地域の可視化画像を図 2 に示す。

次に本モジュールのインターフェースを図6に示す。インターフェースでは時間的条件、空間的条件をユーザが選択することができるようにした。

対象ファイル:	log.txt	選択
現在選択中のファイル:	20081205161000_log.txt	
時間	期間選択	始: 2007年11月25日 終: 2008年12月1日
	粒度(単位)	一日間
地域	地域選択	<input type="radio"/> すべての地域 <input checked="" type="radio"/> 関東地方 <input type="radio"/> 都道府県選択
	粒度(単位)	都道府県まで
条件設定	<input checked="" type="checkbox"/> 有効アクセスのみ	
	<input checked="" type="checkbox"/> 学内アクセスを除く	
	<input checked="" type="checkbox"/> 投稿者を除く	
	<input checked="" type="checkbox"/> 日本国内のみ	
	<input type="checkbox"/> ロボットを除く	
開始ボタン	Google Earth 用ファイル生成	

図6 時間的空間的データベース操作画面

時間的空間的分析モジュールの操作について述べる。まず、IP 地理位置情報が付加されたアクセスログをログ一覧から選択する。その後、ユーザの利用形態に合わせて時間的範囲や粒度、空間的範囲や粒度を選択する。また「出力時に検索エンジンのロボットを除く」などの制約条件（フィルタリングルール）を選択することで、ユーザの様々な要望にこたえることができる。開始ボタンを押すことでKML ファイルの作成を開始する。

6. 地図インターフェースと時間スライダ

本論文では閲覧者の地域を表示する地図インターフェースとして Google 社の「Google Earth」を用いた。Google Earth では XML 形式の KML ファイルを読み込むことができる。試作システムではアクセスログから時間的空間的条件を考慮して閲覧者の地域情報・アクセス時刻情報を反映した KML ファイルを作成した。この作成した KML ファイルを Google Earth に読み込ませることで時間スライダ機能を有効にし、地図上に閲覧者の地域を可視化した(図7, 図8)。

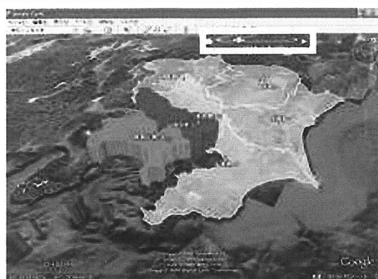


図7 時間スライダ変化前

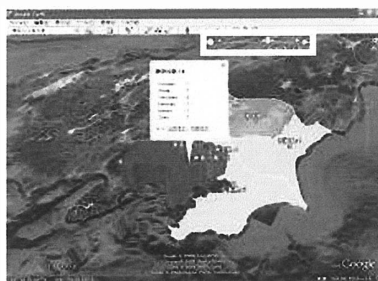


図8 時間スライダ変化後

時間スライダの時間軸を操作することで地図上の閲覧者の地域を表示することができる。時間スライダを操作することで閲覧者の地域の時間的推移が直感的に把握できると考えた。

7. 評価実験

評価実験ではアクセスログにおける閲覧者の地域の直感的把握を目的として試作システム利用に関する実験を行った。

7.1 ユーザの意見

神奈川工科大学公式 Web サイトのアクセスログを変換した KML ファイルを読み込ませた Google Earth を神奈川工科大学企画広報課3名の方に操作してもらい意見をいただいた。アクセスログの集約条件としては有効アクセス(ステータスコードが 200 番)を対象として、学内アクセスを除き、日本国内を対象として、検索エンジンによるアクセスを除いた。また閲覧者のネットワーク地域を都道府県の形状で表現し、ユニークホスト数を色と高さで表現した。試作システムに対する肯定的意見を以下に記す。

- 面白い。地域のアクセスが具体的にわかる
- Google Earth で直感的に把握してから表で詳細を確認するとさらにわかりやすい
- 地方に派遣しているアドバイザーの有効性に

ついて確認できる
また改善点として以下の意見があった。

- 日本全国を対象にすると極端に白と赤にわかれてしまうため、(地域的範囲として日本全国を対象とした場合)アクセス数が少ない地域の差がわかりにくい
- アクセスログを得てから分析を行うため、成果を把握するのに期間がかかる
- 1月 はセンター試験の時期なので、粒度を一週間に設定して試験前と試験後の地域のアクセスの分布を把握したい
- 昼は保護者が Web サイトを閲覧している可能性がある。また夜は高校生が閲覧している可能性がある。そのため表示の時間を朝・昼・夜にしたい
- 粒度を市町村まで対応してほしい
- 地域をクリックしたらその都道府県における市町村のアクセス数を把握したい
- 大学の Web サイトにきた閲覧者のリンク元がわかれば予備校などから貼られている画像リンクが効果的かどうか確認できる
- (地域的な粒度に関して) 市町村もしくは高等学校からのアクセスが把握できればどの高校から注目されていて、逆にどの高校から注目されていないか確認ができる

7.2 考察

評価実験により、本システムはアクセスログから閲覧者の地域を時間的かつ直感的に把握するのに有効と思われる。改善点として時間的粒度や空間的粒度の改善が求められる意見もあり、改善することでシステムの利用範囲が広がると思われる。

8. おわりに

本論文ではアクセス分析における課題を2つ挙げ、IP 地理位置情報を付加したアクセスログの時間的空間的可視化を目的に解決策として試作システムを提案し、評価実験を行った。アクセスログにおける課題として地理位置情報の獲得が困難という課題とアクセス分析ツールの時間的空間的集約操作による直感的な把握は困難という課題を挙げた。

これに対する解決策として、IP 地理位置情報変換を付加したアクセスログの時間的空間的分析システムを提案した。本システムは IP 地理位置情報変換付加モジュールと時間的空間的分析モ

ジュールで構成した。MAXMIND 社の地理位置情報データベースを用いて IP 地理位置情報変換付加モジュールを構築したことでアクセスログの IP 情報を地理位置情報に変換させることが可能になった。時間的空間的分析モジュールでは地理位置情報を付加したアクセスログを関係データベース管理システムで操作することにより、時間的および空間的に集約しインタラクティブに分析が可能となった。また地図インターフェースとして Google 社の GoogleEarth を用いた。評価実験結果より本システムはアクセスログから閲覧者の地域を時間的かつ直感的に把握できることが確認できた。

今後の課題として試作システムの改善が挙げられる。本システムでは日本を対象に地域の粒度を都道府県とした。評価実験ではアクセスログから閲覧者の市町村や高等学校も把握したいというニーズがあった。今後は空間的粒度に関しては市町村まで対応を行い、IP 情報から高等学校などの組織名を得られるように対応したい。本論文が閲覧者の分析の一助となれば幸いである。

参考文献

- 1) 山口, 池端, 伊藤, 梶永, データ宝石箱を用いたウェブアクセスログの視覚化, 第 30 回可視化情報シンポジウム, 2002.
- 2) 梶永, 伊藤, 山口, 池端, ウェブサイトのアクセス分布図を提供する視覚化システム, 情報処理学会マルチメディア通信と分散処理研究, 2002.
- 3) OBAMApPING prototype, <http://rhizome.org/announce/view/51889>
- 4) MaxMind - GeoLite City, <http://www.maxmind.com/app/geolitecity>
- 5) GoogleEarth, <http://earth.google.co.jp>
- 6) 神奈川工科大学, <http://www.kait.jp>