

Blog ベースド地域コミュニティポータル構築

鈴木 秀明, 佐藤 和洋

自治体が運営する地域情報に関する電子会議室や電子掲示板には、荒らしなどが原因で閉鎖に陥った事例が数多く存在している。そこで、本研究では分散コミュニケーションメディアとしての Blog に着目し、議論はシステム内で行うのではなく、ユーザを GIS (Geographical Information System) の地図にリンクされた Blog エントリに誘導する形態を考案した。Blog にはモデレータの役割を担う管理者が存在するため、本システムはコミュニケーションの案内役という立場である。これにより、質の高い情報提供だけでなくシステム管理のコスト削減にもつながり、堅実な運営が可能となる。

地図へのリンクは、Blog の特徴である TrackBack と GPS (Global Positioning System) 携帯端末による Moblog を介して行う。地図を主体とした情報提示により、「what (なに)」の情報検索だけでなく、地図上の詳細な位置を基準とした「where (どこ)」での情報検索も可能になる。

本論文は、こうした考えに基づく Blog ベースド地域コミュニティポータル構築について論じたものである。

1. はじめに

日本における Blog 元年と呼ばれた 2004 年は数十もの Blog サービスサイトが開始された。年が明けて、今年も数多くの Blog 関連サービスが開始されている。特筆すべきは、ポータルサイト最大手 Yahoo! Japan の Blog サービス「Yahoo! ブログ」の開始である。日本におけるインターネット No.1 ページビューを誇る同サイトが Blog サービスを開始したことにより、さらに多くのインターネットユーザに Blog が認知されるのは間違

いなく、2005 年も Blog 熱は冷めそうもない。Blog ユーザが増加するということはより多くの地域情報が生み出され、その集積を目的とする本研究にとっては大きなモチベーションになる。

前回の論文¹では、Blog の特性と動向を踏まえて地域情報を集約するコミュニティポータル構築について述べた。本論文では、そこで掲げた基本コンセプトに基づく Blog ベースド地域コミュニティポータル構築について論じる。

本システムは、江別市を対象とした情報ポータルサイトであり、「Blog コミュニケーションの最大限活用」と「よりローカルな情

SUZUKI Hideaki 2004 年札幌学院大学社会情報学部
佐藤和洋ゼミナール研究生

SATOU Kazuhiro 札幌学院大学社会情報学部

報の蓄積・発信」が目標である。システム内にコミュニケーションスペースを設けるのではなく、地図にリンク（マッピング）された Blog エントリでのコミュニケーションを促進する。Blog エントリのマッピングには、Blog の特徴である TrackBack と GPS 携帯端末による Moblog を介して行う。このほか、基盤情報として江別ブランド事典の各ブランドと地域の商店が運営する Blog のマッピングも行う。

現時点における研究状況は、Web-GIS 機能、TrackBack センター機能、Blog 画像からの位置情報取得機能、RSS サーチサイトからのエントリ取得機能という最低限の機能の実装を終えた。これらはそれぞれ、システムイメージを示した図 1.1 の(A)~(D)にあたる。そして現在は細部の調整・検証・機能追加の段階にある。

論文構成として、第 2 章ではシステムの実装について概説を行い、第 3 章にてシステムの機能説明を述べる。そして第 4 章で考察を述べた後、第 5 章で本研究をまとめる。

2. システムの実装

本章ではシステムの実装について概説する。初めにシステムの環境を紹介し、その上で Web-GIS サーバ機能、「場所」についての

言及を募る TrackBack センター機能、Blog の画像から位置情報を取得してマッピングする機能、RSS サーチサイトからエントリを取得する機能についての実装を述べる。

2.1 システムの環境

システム環境は以下の通りである。

◆ サーバ

OS : Fedora Core² 1

データベース : PostgreSQL³ 7.4.6 + PostGIS⁴ 0.9.1 (CVS)

Web サーバ : Apache⁵ 2.0.50 (Fedora)

GIS サーバ : MapServer⁶ 4.4.1

サーバサイドスクリプト : PHP⁷ 4.3.10 +

PHP MapScript⁸ 4.4.1, Perl⁹ 5.8.6/CGI

GIS フレームワーク : OpenIGF¹⁰ 0.1 Beta

フォント : IPA フォント¹¹

◆ クライアント

Web ブラウザ : Mozilla Firefox¹² 1.0 (Linux, Windows), Internet Explorer 6.0 SP 1 (Windows)

2.2 Web-GIS サーバ機能

本節では、GIS に必須な地図データと、基盤となる「自治体」「郵便番号」「ブランド事典」のリレーションの再設計に関して述べる。

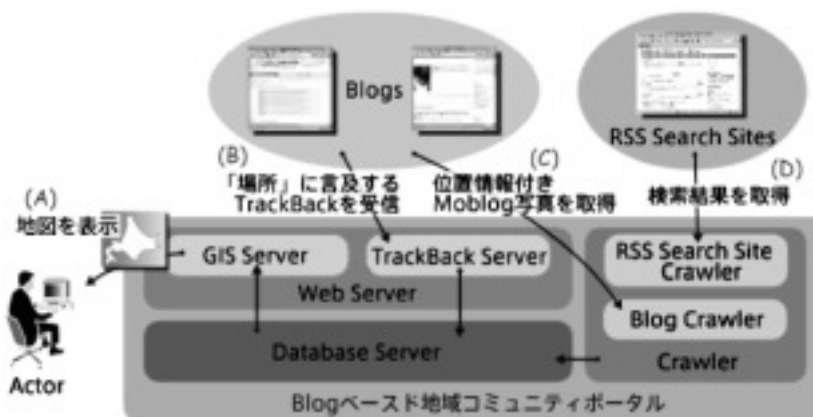


図 1.1 システムイメージ

(1) 地図データ

本システムで使用する地図データとデータベース投入後の E-R 図についての説明を述べる。

◆ 使用地図データ

地図データは、株式会社オークニー¹³の GIS データパック (MapServer 版)¹⁴ 北海道分を使用することにした。これは、国土地理院¹⁵の「数値地図 25000 (空間データ基盤)」、 「数値地図 2500 (空間データ基盤)」および「数値地図 50 m メッシュ (標高)」を基図として作成されたデータ群である。

また、陸地データとして、nishioka 氏が公開する「日本海岸線ポリゴンデータ¹⁶」を用いた。こちらは国土地理院の「数値地図 25000 (行政界・海岸線)」から海岸線データを抽出したポリゴンデータである。

表 2.1 が使用地図データ一覧である。また、以下にデータ形式を示す。

測 地 系：世界測地系 (日本測地系 2000) - JGD2000
 投 影 法：緯度・経度 (10 進度)
 ファイル形式：ESRI Shapefile
 測 量 制 度：数値地図 25000 → 1/25,000

数値地図 2500 → 1/2,500

地図データのデータベース投入は、Shapefile の SQL 変換を行う PostGIS 付属プログラム「shp2pgsql」とその国際化パッチ¹⁷を用いた。

◆ E-R 図

システムの E-R 図を示す (図 2.1, 図 2.2)。図 2.1 が地図データに関するテーブル群である。

テーブルは先述の地図データのほか、OpenGIS¹⁸のメタデータである「spatial_ref_sys」と「geometry_columns」がある。「spatial_ref_sys」は位置情報データベースで使用される ID (SRID) と座標系に関する表記を持つ。「geometry_columns」は位置情報テーブルに関する情報を持つ。

(2) 基盤リレーションの再設計

基盤リレーションの再設計として、各テーブルへの日付情報の付加、郵便番号データの町域名の結合、合併以外で変更された自治体名の修正を行った。

◆ 日付情報の付加

これまでの設計では日付情報がなく、最近の市町村合併による自治体の新設・消滅に対

表 2.1 使用地図データ一覧

ファイル名	内 容	種 別	元データ
coast_pgn.???	陸 地	ポリゴン	数値地図 25000
chimei.???	地 名	ポイント	数値地図 25000
dourokukan.???	道 路 区 間	ラ イ ン	数値地図 25000
gyouseidaihyoten.???	行政代表点	ポイント	数値地図 25000
gyouseikai.???	行 政 界	ラ イ ン	数値地図 25000
kasenkukan.???	河 川	ラ イ ン	数値地図 25000
kokyoshisetsu.???	公 共 施 設	ポイント	数値地図 25000
meshhyoko.???	標 高	ポイント	数値地図 25000
tetsudokukan.???	鉄 道 区 間	ラ イ ン	数値地図 25000
tatemono.???	建 物	ポリゴン	数値地図 2500
zyouti.???	公 園 等 場 地	ポリゴン	数値地図 2500
mizu.???	内 水 面	ポリゴン	数値地図 2500
eki.???	駅 名	ポイント	数値地図 2500

※ファイル名の拡張子???は dbf, prj, shp, shx すべてを指す

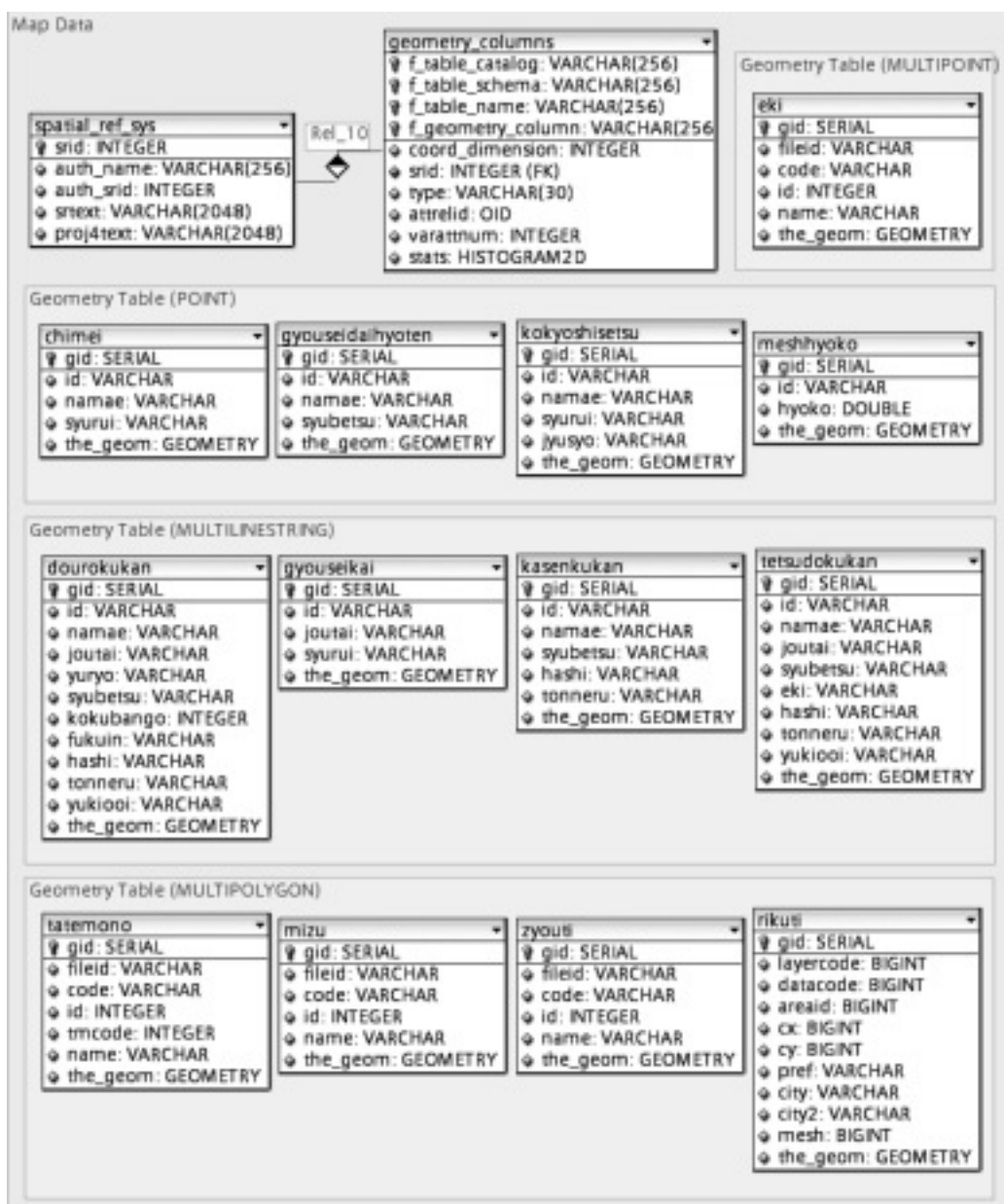


図 2.1 E-R 図 (地図データ)

応することができない。したがって、これらのリレーションを再設計し、日付情報を付加した。自治体に関するデータは、郵便番号データの単体利用から株式会社読売情報開発大阪運営の「ゆうすけ¹⁹⁾」が公開する「基準日時点の市町村データ」と「基準日時点の過去（廃止）市町村データ」²⁰⁾の利用に切り替えること

にした。当データはバッチ処理利用に適した記述形式になっており、自治体の新設や廃止に関わるデータ更新に適している。

◆ 町域名の結合

郵便番号データに関して、京都の地名などで町域名が長いものは複数列に分割されている（図 2.3）ので、自作スクリプト「convert_

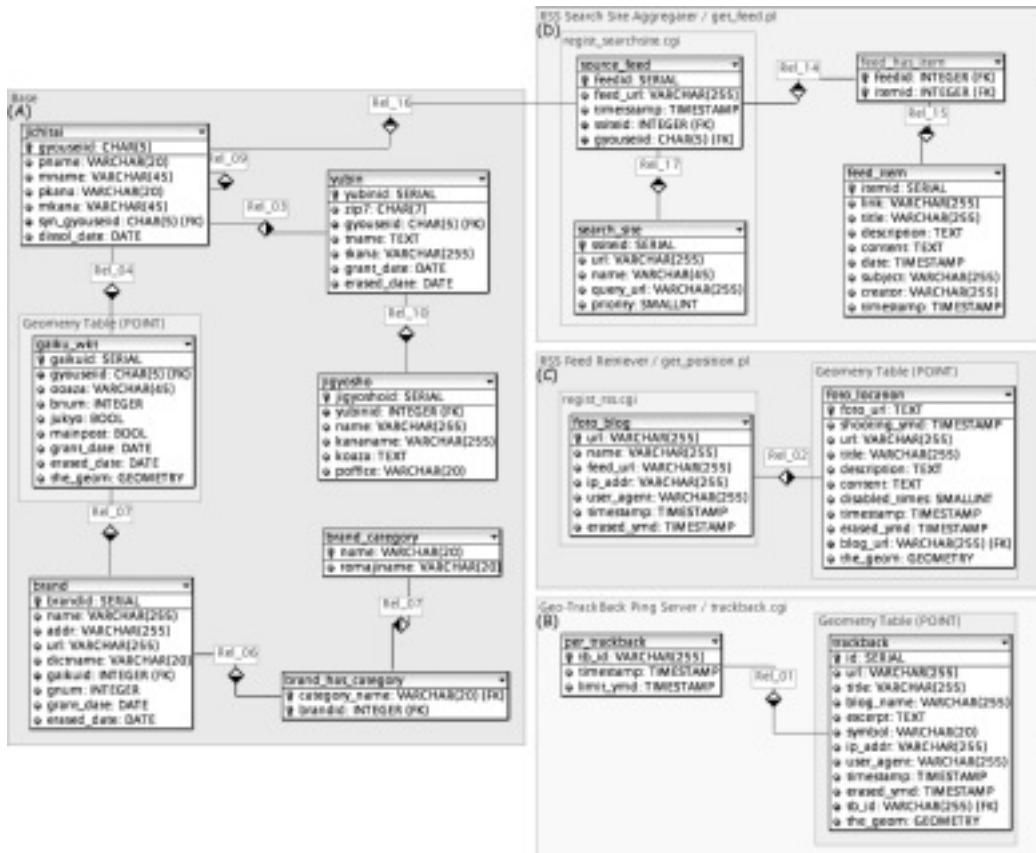


図 2.2 E-R 図 (基盤データ, プログラム取得データ)



図 2.3 郵便番号データ (町域名結合前)

ken.pl」にて1つに統合した(図2.4).

◆ 合併以外で変更された自治体名の修正

2002年4月1日付けで二戸郡安代町が岩手郡安代町に改名²¹されているため、変更されていない事業所データの挿入エラーが発生する。そこで古いデータをテーブル「自治体」に登録する(廃止年月日や新規自治体名も挿入する)。

また、2003年4月1日付けで政令指定都市になったさいたま市²²に関しては、1つのレコードだったのに対して複数の区に関するレコードに分裂したため、テーブルを「多対多」にする必要が生じてしまう。そこで、さいたま市のみ手動入力する(「廃止日」のみで「編

入先市区町村コード」の値は入力しない)。

◆ E-R図

図2.2の(A)が基盤リレーションに関するテーブル群である。

地方公共団体の情報を持つ「jichitai」、郵便番号情報を持つ「yubin」、大口事業所の郵便番号情報を持つ「jigyosho」、街区レベル位置参照情報を持つ「gaiku_wkt」、江別ブランド事典に関する情報を持つ「brand」、江別ブランド事典のカテゴリ情報を持つ「brand_category」とその関連実体の「brand_has_category」がある。

◆ 属性定義表

属性定義表を示す(表2.2~表2.8)。

郵便番号	市区町村	住所
26102,602,6028358	京都府京都市上京区北田	上立売通天神道西入上る、上立売通御前西入上る、上立売通御前西入2丁目、上立売通御前西入2筋目、下長者通御前西入、天神道上立売上る、天神道仁和寺街道下る、天神道下立売上る、天神道妙心寺道上る、天神道妙心寺道下る、仁和寺街道天神道西入下る、仁和寺街道天神道東入下る、御前通上立売上る、御前通上立売上る西入、御前通下立売上る、御前通下長者通上る西入、御前通仁和寺街道下る西入、御前通妙心寺道上る西入、御前通西裏上立売上る、御前通西裏下立売上る
26102,602,6028392	京都府京都市上京区北田	御前通五辻上る、御前通今出川上る、御前通今出川上る1丁目、御前通今出川上る2丁目、御前通今出川上る2丁目西入、御前通今出川上る2筋目、御前通今出川上る西入、御前通今出川上る東入、御前通今出川2丁目上る、御前通寺之内下る
26102,602,6020818	京都府京都市上京区北横田	
26102,602,6020027	京都府京都市上京区木下突抜町	
26102,602,6028146	京都府京都市上京区木屋之町	
26102,602,6020851	京都府京都市上京区九軒町	
26102,602,6020881	京都府京都市上京区京都御苑	

図2.4 郵便番号データ(町域名結合後)

表2.2 jichitai

物理属性名	データ型	PK	FK	NN	論理属性名
gyouseiid	CHAR (5)	Yes		Yes	地方公共団体ID
pname	VARCHAR (20)			Yes	都道府県名
mname	VARCHAR (45)			Yes	市区町村名
pkana	VARCHAR (20)				都道府県名カナ
mkana	VARCHAR (45)				市区町村名カナ
syn_gyouseiid	CHAR (5)		Yes	Yes	統合先自治体ID
dissol_date	DATE				廃止日

表 2.3 yubin

物理属性名	データ型	PK	FK	NN	論理属性名
yubinid	SERIAL	Yes		Yes	郵便 ID
zip7	CHAR (7)			Yes	郵便番号
gyouseiid	CHAR (5)		Yes	Yes	地方公共団体 ID
tname	TEXT				町域名
tkana	VARCHAR (255)				町域名カナ
grant_date	DATE				作成日
erased_date	DATE				削除日

表 2.4 jigyocho

物理属性名	データ型	PK	FK	NN	論理属性名
jigyoshoid	SERIAL	Yes		Yes	大口事業所 ID
yubinid	INTEGER		Yes	Yes	郵便 ID
name	VARCHAR (255)			Yes	大口事業所名
kananame	VARCHAR (255)			Yes	大口事業所名カナ
koaza	TEXT			Yes	小字名・丁目・番地等
poffice	VARCHAR (20)			Yes	取扱郵便局名

表 2.5 gaiku_wkt

物理属性名	データ型	PK	FK	NN	論理属性名
gaikuid	SERIAL	Yes		Yes	街区 ID
gyouseiid	CHAR (5)		Yes	Yes	地方公共団体 ID
ooaza	VARCHAR (45)				大字・町丁目
bnum	INTEGER			Yes	街区符号・地番
jukyo	BOOL			Yes	住居表示フラグ
mainpost	BOOL			Yes	代表フラグ
grant_date	DATE				作成日
erased_date	DATE				削除日
the_geom	GEOMETRY			Yes	緯度・経度

表 2.6 brand

物理属性名	データ型	PK	FK	NN	論理属性名
brandid	SERIAL	Yes		Yes	ブランド ID
name	VARCHAR (255)			Yes	ブランド名
addr	VARCHAR (255)			Yes	ブランド住所
url	VARCHAR (255)			Yes	ブランド URL
dictname	VARCHAR (20)			Yes	ブランド事典名
gaikuid	INTEGER		Yes	Yes	街区 ID
gnum	INTEGER				号・枝番以降
grant_date	DATE			Yes	作成日
erased_date	DATE				削除日

表 2.7 brand_category

物理属性名	データ型	PK	FK	NN	論理属性名
name	VARCHAR (20)	Yes		Yes	カテゴリ名
romajiname	VARCHAR (20)			Yes	カテゴリ名ローマ字

表 2.8 brand_has_category

物理属性名	データ型	PK	FK	NN	論理属性名
category_name	VARCHAR (20)	Yes	Yes	Yes	カテゴリ名
brandid	INTEGER	Yes	Yes	Yes	ブランド ID

2.3 「場所」についての言及を募る TrackBack センター機能

本節では、Geo-TrackBack Ping Server とそれに関するプログラム、そしてデータベース構成についての概説を行う。

(1) Geo-TrackBack Ping Server

本プログラムは、「場所」についての言及を募る TrackBack サーバを実現するためのプログラムである。場所の識別には緯度・経度 (lat/long) を用いる。現時点では位置情報を対象にした Blog の検索エンジンは存在しておらず、その意味で、位置情報をキーにした情報を集約すれば地域情報ポータル of キーコンテンツとなり得ると考える。緯度・経度だけでなく、一般的な TrackBack ID を用いた TrackBack も受け付ける。位置情報を基にした TrackBack の利用に際する流れを図 2.5 に示す。



図 2.5 Geo-TrackBack Ping Server のイメージ

(A) TrackBack Ping URL を取得する

TrackBack を送信したい地点を表示し、マッピング用アイコンを選択するとパラメータに緯度・経度およびマッピング用アイコン名が含まれる TrackBack Ping URL が表示される。(例: <http://example.com/cgi-bin/trackback.cgi/E141.49559/N43.06594/BLUE>, 「E141.49559」が経度, 「N43.06594」が緯度, 「BLUE」がマッピング用アイコン名を示す)

(B) TrackBack Ping を送信する

表示された TrackBack Ping URL に自分の Blog から TrackBack Ping を送信する。

(C) エントリをマッピングする

パラメータの位置情報を基に、送信元エントリをマッピングする。システムは、データベースに問い合わせ、地図データが存在する場所のみ TrackBack を受け付ける。TrackBack が成功するとテーブル「trackback」に位置情報、マッピング用アイコン名、エントリの URL、投稿日時などを格納する。

TrackBack ID を用いた TrackBack の場合はパラメータに「tbeetsu」のように 1 語を指定する。この際、テーブル「per_trackback」の属性「tb_id」に登録されている TrackBack

ID についてのみ、TrackBack を受け付ける。TrackBack が成功するとテーブル「trackback」に情報が保存される。

(2) プログラム

Geo-TrackBack Ping Server に関するプログラムの概説を行う。位置情報を基にした TrackBack を受け付ける「trackback.cgi」がある。また、TrackBack プロトコルの脆弱性についても述べる。

◆ trackback.cgi

本プログラムは、Blog から TrackBack Ping を受信し、データベースへの登録・削除を行うプログラムである。

➤ 情報の登録

緯度・経度が含まれたパラメータを受け取ると、緯度・経度・マッピング用アイコン名の3つに分割する。取得した緯度・経度がシステムの地図データ範囲内のものであるかを問い合わせ、範囲内であることが確認されると TrackBack 元 URL などとともにデータベースに登録する。

地図データ範囲の判定は PostGIS の extent 関数で矩形を生成し、その範囲に沿った処理を行う。ただし、extent 関数が生成する地形は矩形であるため、かなり大まかな範囲であり、正確な境界線ではない。この代わりに Polygonize 関数を用いて正確な境界ポリゴンを作成することも可能である。

➤ 情報の削除

登録時と同様（近隣）の地点に再び TrackBack を行うことで削除が可能である。

◆ TrackBack サーバ脆弱性

多くの TrackBack サーバの実装には、TrackBack スパムの標的になりうる脆弱性が指摘されている。問題点を洗い出し、その対処法を示す。

➤ 問題点

悪意のあるサイトが JavaScript を使用し、訪問者にスパム Ping の送信を強制させるこ

とが可能のため、クライアントの IP アドレスをはじく手段では対応できない。この脆弱性を XSS（クロスサイトスクリプティング）脆弱性と組み合わせると事態はより深刻となる。この脆弱性の根本的理由は、TrackBack は一般的な Web ブラウザから POST メソッドで送信でき、また TrackBack 送信元のなりすまし防止が不可能なプロトコルということにある。

➤ 対処法

まず悪意あるスパマーによる TrackBack の連続投稿を防止するために、同一 IP アドレスからの連続 TrackBack を排除するようにした。そして、JavaScript を用いた TrackBack 送信ツールを悪用したスパムを防止するために、一般的な Web ブラウザからの発信を排除する。具体的には、User Agent の値に「Mozilla/」が含まれているときや HTTP Referer に値が存在するときである。このほかの手段としては、多くのスパムは英語などの1バイト文字であることから、マルチバイト文字以外は受け付けないようにする方法や、ひらがなの含まれない TrackBack を排除するといった方法もある。

(3) データベース構成

データベースの構成について概説する。はじめに E-R 図について説明し、続いて属性定義表について説明する。

◆ E-R 図

図 2.2 の(B)が TrackBack サーバ機能に関するテーブル群である。

TrackBack ID を用いた TrackBack の受信時に利用する「per_trackback」と、受信した TrackBack の情報を持つ「trackback」がある。

◆ 属性定義表

属性定義表を示す（表 2.9、表 2.10）。

表 2.9 per_trackback

物理属性名	データ型	PK	FK	NN	論理属性名
tb_id	VARCHAR (255)	Yes		Yes	TrackBack テーマ ID
timestamp	TIMESTAMP			Yes	登録日時
limit_ymd	TIMESTAMP				投稿期限日時

表 2.10 trackback

物理属性名	データ型	PK	FK	NN	論理属性名
id	SERIAL	Yes		Yes	TrackBack ID
url	VARCHAR (255)			Yes	TrackBack 元 URL
title	VARCHAR (255)			Yes	TrackBack 元タイトル
blog_name	VARCHAR (255)			Yes	TrackBack 元 Blog 名
excerpt	TEXT				TrackBack 元エントリ引用
symbol	VARCHAR (20)				マッピング用アイコン名
ip_addr	VARCHAR (255)			Yes	投稿者 IP アドレス
user_agent	VARCHAR (255)				投稿者ユーザーエージェント
timestamp	TIMESTAMP			Yes	投稿日時
erased_ymd	TIMESTAMP				削除日時
tb_id	VARCHAR (255)		Yes		TrackBack テーマ ID
the_geom	GEOMETRY				緯度・経度

2.4 Blog の画像から位置情報を取得して地図とリンクする機能

本節では、RSS Feed Retriever とそれに関するプログラム、そしてデータベース構成についての概説を行う。

(1) RSS Feed Retriever

本プログラムは、Blog の画像から位置情報を取得して地図とリンクするクローラである。先述の TrackBack センター機能は、Blog の管理画面から TrackBack Ping を送信する必要があるため、対象は PC であり Moblog には対応しない。本機能は GPS 対応携帯端末による Moblog を対象とする。システムが対象の Blog に投稿された画像データから位置情報を取得し、そのエントリを地図と結びつける機能がある。利用に際する流れを図 2.6 に示す。



図 2.6 RSS Feed Retriever のイメージ

(A) 対象 Blog の RSS フィードを登録する

RSS フィードを登録する方法は、あらかじめ自分の Blog のエントリに「ポータルに登録したいです」といった合言葉を記入しておく。その上で、RSS フィード URL

か対象 Blog 自体の URL をシステムに登録する。Blog 自体の URL を指定した場合は、システム側が RSS Auto Discover を利用して RSS フィード URL を抽出する。するとすぐにシステム側から対象 Blog のチェックを行う。合言葉が対象 Blog に記述されていることを確認するとデータベースにデータを登録する。

(B) RSS フィードを定期的に更新チェックする

システムに登録した RSS フィードを定期的にチェックする。

(C) エントリから JPEG 画像を取得する

更新されたエントリがあり、なおかつ JPEG 画像があればその画像を取得する。ただし、エントリ中に「<stay away>」という文字列が含まれていた場合にはそのエントリの画像を取得しないものとする。RSS フィードをシステムに登録すると、以降は常に地図に関連付けられることになるが、これを一時的に拒否するための仕組みである。なお、「<>」は HTML タグを示し、本来ブラウザ上には表示されない。ただし、中には Moblog 後に実体参照化する Blog も存在するため「<stay away>」も対象とする。

(D) EXIF から位置情報を抽出する

JPEG 画像のメタデータである EXIF に位置情報があればこれを抽出する。なお、1つのエントリからは1つの位置情報画像が取得できればよい。これは、位置情報を複数取得してもたいていは近場で撮影した写真である可能性が高いと予測できるためである。近場に同じエントリが複数マッピングされることになればそれは冗長である。ここで重要なのは、本システムは「写真」単位でマッピングするのではなく「エ

ントリ」単位でマッピングするという点である。

(E) エントリをマッピングする

抽出した位置情報を基に、取得エントリをマッピングする。位置情報、エントリの URL、投稿日時などをデータベースに格納する。

(2) プログラム

RSS Feed Retriever に関するプログラムの概説を行う。プログラムは登録・巡回処理をするスクリプト「get_position.pl」と RSS フィードの引数を受け取りスクリプトに受け渡す CGI「regist_rss.cgi」に分かれる。

◆ regist_rss.cgi

本プログラムは、システムに GPS エントリ取得用 RSS フィード URL の登録・削除を行うプログラムである。

➤ 情報の登録

CGI の「rss」パラメータで RSS フィードの値を受け取ると、この RSS フィードを取得・解析する。RSS フィード中にあらかじめ指定した「合言葉」が記述されているとデータベースに対象 Blog の情報を登録する。取得・解析に失敗したり、既に登録されている場合はエラーを表示する。

➤ 情報の削除

削除も登録の手順と同様である。

◆ get_position.pl

本プログラムは、regist_rss.cgi により登録された Blog 情報を抽出し、対象 Blog の RSS フィードの中の description エlement もしくは content:encord エlement に含まれる画像を取得する。さらに、その画像の EXIF 情報を解析し、位置情報が含まれていた場合はそのエントリ情報をデータベースに登録する。

情報の取得は cron デーモンを利用して定期的に行うが、規定回数の取得に失敗した場合は、該当 Blog に関するデータを全て消去

する。対象 Blog への負荷やトラフィックなどの問題を考慮し、キャッシュ機能も実装している。

(3) データベース構成

データベースの構成について概説する。はじめに E-R 図について説明し、続いて属性定義表について説明する。

◆ E-R 図

図 2.2 の(C)が GPS 写真を含むエントリ (GPS エントリ) の取得機能に関するテーブル群である。

巡回対象の Blog に関する情報を持つ「foto_blog」と、GPS エントリの情報を持つ「foto_location」がある。

◆ 属性定義表

属性定義表を示す (表 2.11, 表 2.12)。

2.5 RSS サーチサイトからエントリを取得する機能

本節では、RSS Search Site Aggregator とそれに関するプログラム、そしてデータベース構成についての概説を行う。

(1) RSS Search Site Aggregator

本プログラムは、複数の RSS サーチサイトのデータを統合し、1つにまとめることで特定の情報を網羅することを目的としたクローラである。利用に際する流れを図 2.7 に示す。

(A) 地域情報の Push 型検索を行う

管理者はあらかじめ各 RSS サーチサイトで地域情報 (例: 「江別 OR 野幌 OR おおさき」) を検索しておき、その検索結果の RSS フィードをシステムに登録する。

表 2.11 foto_blog

物理属性名	データ型	PK	FK	NN	論理属性名
url	VARCHAR (255)	Yes		Yes	Blog URL
name	VARCHAR (255)				Blog 名
feed_url	VARCHAR (255)			Yes	RSS フィード URL
ip_addr	VARCHAR (255)			Yes	登録者 IP アドレス
user_agent	VARCHAR (255)				登録者ユーザーエージェント
timestamp	TIMESTAMP			Yes	登録日時
erased_ymd	TIMESTAMP				削除日時

表 2.12 foto_location

物理属性名	データ型	PK	FK	NN	論理属性名
foto_url	TEXT	Yes		Yes	写真 URL
shooting_ymd	TIMESTAMP			Yes	撮影日時
url	VARCHAR (255)			Yes	エントリ URL
title	VARCHAR (255)				エントリタイトル
description	TEXT				エントリ要約
content	TEXT				エントリ全文
disabled_times	SMALLINT			Yes	取得失敗回数
timestamp	TIMESTAMP			Yes	登録日時
erased_ymd	TIMESTAMP				削除日時
blog_url	VARCHAR (255)		Yes	Yes	Blog URL
the_geom	GEOMETRY			Yes	緯度・経度



図 2.7 RSS Search Site Aggregator のイメージ

なお、Push 型検索とは、検索結果の RSS フィードを定期的に調べることにより、常に最新の情報が取得可能な検索形態と定義する。つまり、Push 型検索は検索結果を RSS フィードとして配信するサーチサイトである必要がある。このような検索を「未来検索」と表記する場合もある。

(B) 検索結果の RSS フィードを集約

システムに登録した検索結果の RSS フィードを定期的に巡回・取得する。サーチサイトの検索結果を統合することによ

て情報を網羅的に扱うことができるが、そのほとんどは重複データであることが予想される。そこで、データが重複した際の優先利用順位をつけ、エントリの URL をキーとして 1 つにまとめる。

(C) 集約したエントリリストを RSS フィードとして配信

データベースに集約したエントリリストはサイトに直接表示するとともに、RSS フィードとしても配信する。

(B) の RSS サーチサイトの優先利用順位の決定は、各 RSS サーチサイトの特性を分析し、体裁などがどれだけオリジナルに近く再現されているかを判断基準とした。本機能の目的はエントリへのリンクを抽出することが目的である。その為に各サーチサイトの統合を行う故、インデックス数といった検索性能そのものの比較評価はしない。

比較対象は Blog の RSS フィードを用いた検索が可能で、検索結果を RSS フィードで配信可能である、という 2 点を備えたサイトである。表 2.13 がこれを満たす 10 個のサイトを比較した表である。要点とな

表 2.13 RSS サーチサイトの比較

サイト名	URL	RSS バージョン	content:encode サポート	description 全文再現	HTML 改行	HTML アンカー	備考
Bulkfeeds	http://bulkfeeds.net/	1.0	○	○	○	○	
FeedBack	http://naoya.dyndns.org/feedback/	1.0	×	×	○	○	一部機種依存文字の削除
未来検索 livedoor	http://sf.livedoor.com/	1.0	×	×	×	×	英数字・記号を半角に、機種依存文字を「?」に置換 スペースの削除
blogdb.jp	http://blogdb.jp/	1.0	○	○	○	○	
goo BLOG	http://blog.goo.ne.jp/	1.0	×	×	×	×	キーワード周辺の主体表示 一部特殊文字の削除
Blogzine	http://blog.ocn.ne.jp/	1.0	×	×	×	×	goo BLOG のエンジンを使用
Exblog	http://www.exblog.jp/	2.0	×	×	○	×	キーワード周辺の主体表示
BlogPeople	http://blogpeople.net/	1.0	×	○	×	×	スペースの削除
もふろげっと	http://mobloget.jp/	2.0	×	×	×	×	RSS フィードではなく、ページ全体の文字列が対象
NAMAAN	http://www.namaan.net/	1.0	×	×	×	×	先頭 25 文字でカット

る項目の説明を下記に示す。

◆ content:encode サポート

サーチサイトで配信される RSS フィードにおいて、item セクションの content:encode エレメントをサポートしているかを表す項目である。なお、content:encode エレメントにはエントリのフルコンテンツが記述される。

◆ description 全文再現

サーチサイトで配信される RSS フィードにおいて、item セクション内の description エレメントが配信元 Blog のオリジナルと一致するかを表す項目である。「○」は文章がすべて再現されているもの、「×」は一部削除されているものを示す。なお、description エレメントには一般にエントリの要約が記述される。

◆ HTML 改行

description エレメントもしくは content:encode エレメントのいずれかにおいて、HTML の改行をサポートしているかを表す項目である。「○」はサポートされているもの、「×」はサイトによって削除されることを示す。

◆ HTML アンカー

description エレメントもしくは content:encode エレメントのいずれかにおいて、HTML のアンカーをサポートしているかを表す項目である。「○」はサポートされているもの、「×」はサイトによって削除されることを示す。

◆ 備考

以上の項目の補足説明を示す。

これらの情報から、RSS サーチサイトの優先順位を以下の通りに設定する。ただし、RSS フィードの 2 次利用に際して何らかの問題が発生する可能性もあるため、実際の対外的な運用に際しては各サイトにおける著作権の確認が必須となる。

1. Bulkfeeds

2. blogdb.jp
3. BlogPeople
4. FeedBack
5. Exblog
6. 未来検索 livedoor
7. NAMAAN
8. goo BLOG
9. Blogzine
10. もぶろげっと

(2) プログラム

プログラムは登録・巡回処理をするスクリプト「get_feed.pl」と RSS フィードの引数を受け取りスクリプトに受け渡す CGI 「regist_searchsite.cgi」に分かれる。

◆ regist_searchsite.cgi

本プログラムは、システムに RSS サーチサイトとその検索用 RSS フィード URL の登録・編集・削除を行うプログラムである。

➤ 情報の登録

HTML フォームから RSS サーチサイト URL と優先度の値を受け取ると、この RSS サーチサイト URL の有効確認を行う。有効を確認した場合はテーブル「search_site」にそのサイト名、RSS サーチサイト URL、優先度を登録する。続いて、この情報を基に検索用 RSS フィード URL を登録する。ユーザが HTML フォームに市町村名と検索ワードを指定すると自動的に検索用 RSS フィード URL を生成し、データベースに登録する。取得・解析に失敗したり、既に登録されている場合はエラーを表示する。

➤ 情報の修正

RSS サーチサイト URL リストに表示される「編集」を選択して RSS サーチサイト URL、優先度、タイトルを編集することができる。検索用 RSS フィード URL の修正はできない。

➤ 情報の削除

RSS サーチサイト URL リストに表示さ

れる「削除」を選択することで削除可能である。

◆ get_feed.pl

本プログラムは、regist_searchsite.cgiにより登録された検索用RSSフィードURLを優先度に従って抽出し、検索結果のRSSフィードを取得する。この中からエントリ情報を抽出してデータベースに登録する。対象Blogへの負荷やトラフィックなどの問題を考慮し、キャッシュ機能も実装している。定期的な取得にはcronデーモンを利用する。

(3) データベース構成

データベースの構成について概説する。はじめにE-R図について説明し、続いて属性定義表について説明する。

◆ E-R図

図2.2の(D)がRSSサーチサイトの結果取得機能に関するテーブル群である。

RSSサーチサイトの情報を持つ「search_site」のほか、検索結果RSSフィードの情報を持つ「source_feed」、取得したエントリ情報を持つ「feed_item」とその関連実体の「feed_has_item」がある。「source_feed」は、基盤リリースの「jichitai」と「多対一」の関係

表 2.14 search_site

物理属性名	データ型	PK	FK	NN	論理属性名
ssiteid	SERIAL	Yes		Yes	サーチサイト ID
url	VARCHAR (255)			Yes	サーチサイト URL
name	VARCHAR (45)			Yes	サーチサイト名
query_url	VARCHAR (255)			Yes	サーチサイトクエリ URL
priority	SMALLINT			Yes	優先度

表 2.15 source_feed

物理属性名	データ型	PK	FK	NN	論理属性名
feedid	SERIAL	Yes		Yes	フィード ID
feed_url	VARCHAR (255)			Yes	RSS フィード ID
timestamp	TIMESTAMP			Yes	登録日時
ssiteid	INTEGER		Yes	Yes	サーチサイト ID
gyouseiid	CHAR (5)		Yes		地方公共団体 ID

表 2.16 feed_item

物理属性名	データ型	PK	FK	NN	論理属性名
itemid	SERIAL	Yes		Yes	アイテム ID
link	VARCHAR (255)			Yes	エントリ URL
title	VARCHAR (255)			Yes	エントリタイトル
description	TEXT			Yes	エントリ要約
content	TEXT				エントリ全文
date	TIMESTAMP			Yes	エントリ投稿日時
subject	VARCHAR (255)				カテゴリ名
creator	VARCHAR (255)				投稿者
timestamp	TIMESTAMP			Yes	投稿日時

表 2.17 feed_has_item

物理属性名	データ型	PK	FK	NN	論理属性名
feedid	INTEGER	Yes	Yes	Yes	フィード ID
itemid	INTEGER	Yes	Yes	Yes	アイテム ID

がある。

◆ 属性定義表

属性定義表を示す (表 2.14~表 2.17)。

3. システムの機能説明

本章では、実際の利用イメージを示してシステムの機能説明を行う。

◆ 初期画面

システムの初期画面を図 3.1 に示す。初期表示地点は北海道江別市全域としている。画面はフレームで分割されており、左ペインに地図を表示する。各ツールによる問い合わせはフレームの右ペインに表示され、下ペインでは縮尺の変更や使用中ツール、表示中心地点の住所を表示する。ツールバーには「ズームイン」「ズームアウト」「移動」「初期画面」「トラックバックの検索」「トラックバック URL の表示」「距離の計測」がある。

◆ 凡例

メニューの「凡例」をクリックすると、右ペインに凡例を表示する (図 3.2)。

◆ レイヤ

メニューの「レイヤ」をクリックすると、

右ペインにレイヤー一覧を表示する (図 3.3)。

レイヤ名左のチェックボックスを切り替えて「地図の再読込」をクリックするとレイヤの表示・非表示を切り替えることができる。

表示可能レイヤは、「トラックバック」「陸地」「ブランド」「場地」「大型河川・湖沼」「建物」「河川」「行政界」「無料道路」「有料道路」「鉄道」「駅」「行政代表点」「公共施設」「地名」である。

◆ 地域のブログ

メニューの「地域のブログ」をクリックすると、右ペインに表示中の地域に関する Blog 一覧を表示する (図 3.4)。地図の表示地点を変更したときに「表示地点の更新」をクリックすると、該当地域の Blog 一覧に更新する。「詳細表示」をクリックすると、エントリの要約を含めた詳細を新規ウィンドウに表示する。「RSS フィード」リンクで RSS フィードを配信する。

データは RSS Search Site Aggregator により取得されるが、現在は江別市のデータのみに対応となっている。



図 3.1 初期画面



図 3.2 凡例

◆ ブログの登録

メニューの「ブログの登録」をクリックすると、右ペインに GPS 画像を配信する Blog を登録することができる (図 3.5).

データは RSS Feed Retriever により取得される。

◆ トラックバック URL の表示

ツールメニューの「トラックバック URL の表示」アイコンを選択した上で表示させたい地点をクリックすると、右ペインに TrackBack Ping URL 表示用の画面を表示する (図 3.6). マッピング時に表示させるフラッグの色を選択すると、その下にあるテキストボックスに TrackBack Ping URL を表示する。TrackBack の完了後に「地図の再読み込み」をクリックすると地図に反映される。

データは Geo-TrackBack Ping Server により取得される。

◆ トラックバックの検索

ツールメニューの「トラックバックの検索」アイコンを選択した上で検索したい地点をドラッグで選択すると、右ペインに選択地点に関する TrackBack 一覧を表示する (図 3.7).

◆ ブランドの検索

ツールメニューの「ブランドの検索」アイコンを選択した上で検索したい地点をドラッグで選択すると、右ペインに選択地点に関するブランド一覧を表示する (図 3.8).

4. 考察

本研究では、「Blog コミュニケーションの



図 3.3 レイヤ



図 3.4 地域のブログ



図 3.5 ブログの登録



図 3.6 トラックバック URL の表示



図3.7 トラックバックの検索



図3.8 ブランドの検索

最大限活用」と「よりローカルな情報の蓄積・発信」を目指してシステムの開発を行ってきた。本章では以上の流れを踏まえ、Webメディアを介したコミュニケーション、地域情報の収集・蓄積、システムの実装という3つの切り口で考察を述べる。

4.1 Webメディアを介したコミュニケーションについて

自治体がオフィシャルで情報ポータルサイトを運営するには、障害となる問題がいくつも存在し、難しさがある。自治体が採用する代表的なWebメディアには、電子会議室や電子掲示板がある。しかし、荒らし・フレーミングによる閉鎖や、反対に書き込みがないなどの原因で閉鎖に陥った事例も数多く存在しており、北海道では夕張市、栗山町、北広島市がその一例である。成功例として有名な神奈川県大和市や同藤沢市、三重県は、学術機関などと連携した大々の組織での運用を行っており、他自治体で同様に実践するにはコストがかかりすぎる。こうした事例があるため、江別ブランド事典は市民から募る情報をまちの資源の中で良いと思える事物、つまり「ブランド」に絞り、さらに市の監査を通すことを考えたと予想できる。ただし、こうした形態のため江別ブランド事典はコミュニケーションメディアではない。

そこで、本研究では分散コミュニケーションメディアとしてのBlogに着目し、議論はシステム内で行うのではなく、ユーザをリンク先のBlogに誘導する形態を考案した。本システムでは、ユーザはマッピングされた情報からリンクをたどって、江別ブランド事典や各Blogへの移動が可能である。江別ブランド事典では、より詳細な情報と江別市民による評価を閲覧することができ、リンク先のBlogでは、コメント機能などを利用してBlogの管理者とコミュニケーションを図ることができる。Blogにはモデレータの役割を担う管理者が存在するため、本システムはコミュニケーションの案内役という立場である。マッピングされたBlogエントリーには、その地点に関する話題が掲載されているため、明確な議論のテーマも存在する。こうした形態の採用により、質の高い情報提供だけでなくシステム管理のコスト削減にもつながり、堅実な運営が可能となる。

4.2 地域情報の収集・蓄積について

まずは地図のレイヤとして江別ブランド事典のデータを組み込むことにしたが、こういった自治体主導で収集された信頼できる情報源を基盤情報の一部として用いることにより、システムの質の向上に役立つと考えた。また、江別ブランド事典側としても地図と結

びつけることができるため、今までは不可能だった店舗などの位置確認に役立つといったメリットもある。その上で、TrackBack と GPS 写真による位置情報蓄積のアプローチと RSS サーチサイトを用いて、地域住民による細やかかつ多彩な Blog エントリを取得する。これにより、「what (なに)」の情報検索だけではなく、地図上の詳細な位置を基準とした「where (どこ)」での情報検索も可能になる。そして、情報の更新頻度が低い江別ブランド事典の補完という役割も持つ。

このように、多くはない特定地域の様々な角度からの情報抽出と、情報配信元 (Blog) でのコミュニケーションという地域コミュニティポータルの骨組みができた。ただし、当初の目的であった個人商店が運用する Blog のマッピングが実現していないため、TrackBack や GPS エントリとは違う形態でのマッピングの仕組みを考える必要がある。これが実現に至り、各 Blog と連携してクーポン情報などを提供することができるになれば、地域に根ざした情報がより安定的に集約されることになるだろう。

なお、江別ブランド事典を開発する江別経済ネットワーク²³では、江別ブランド事典として開発したプログラムのオープンソース化を表明²⁴している。このため、他自治体でのブランド事典開設の際には江別ブランド事典と同様の手段で対応できると考えられる。

4.3 システムの実装について

プログラムを作成する際に感じたが、Perl や PHP にはモジュール類が豊富にそろっており、それらを利用することで実装期間の短縮が可能となった。ただ、画像から位置情報を抽出するモジュールでは、多少柔軟性に欠く部分があり、その解決法を見いだすまで多少時間を費やした。また、RSS フィードをパースするモジュールが EUC-JP や Shift_JIS といった日本語エンコードに対応してい

ないため、うまく処理できないという問題にも遭遇した。そのほか、既存プログラムを参考にしたものもあるが、明らかなバグがあり結局全て書き直すという事もあった。

位置情報を扱う際には、測地系や投影法に関しての調査不足のために解決まで時間がかかった。例えば、EXIF の位置情報は、GPS 端末によって「35/1, 39/1, 3010/100」という「度分秒(DMS)」形式だけでなく、「3568712/100000, 0/1, 0/1」といった「度(D)」のみ(10進度)で表記する形式があるため注意が必要だった。

5. おわりに

本論文では、Blog ベース地域コミュニティポータルの実装について論じた。位置情報をベースにした TrackBack センター機能を実現する Geo-TrackBack Ping Server と RSS フィードから位置情報付き画像を抜き出す RSS Feed Retriever について概説し、複数の RSS サーチサイトから抽出した地域情報を統合する RSS Search Site Aggregator の概説も行った。その上で、全体の考察も述べた。

現在は基本的な機能の実装を終えた段階にある。次の展開として、機能向上のほか、パフォーマンス問題やユーザビリティの向上、住所データ・地図データ陳腐化への対応、RSS サーチサイトの著作権確認、プライバシー問題の注意喚起など、改善・解決すべき点の対応を行う。

実際のシステム運用には多くの情報蓄積が必要となる。そのためにはコミュニティの獲得が重要となるため、幅広い協力の呼びかけも行っていきたい。

参考文献

- 1 鈴木秀明, 佐藤和洋: Blog 技術の動向と Blog ベース地域コミュニティポータルの基本構想, 社会情報, Vol.14 No.1, pp.61-86 (Dec. 2004)

- 2 Fedora Project: <http://fedora.redhat.com/>
- 3 PostgreSQL: <http://www.postgresql.org/>
- 4 PostGIS: <http://postgis.refractor.net/>
- 5 The Apache HTTP Server Project: <http://httpd.apache.org/>
- 6 MapServer Homepage: <http://mapserver.gis.umn.edu/>
- 7 PHP: Hypertext Preprocessor: <http://www.php.net/>
- 8 PHPMapScript.MapTools.org: http://www.maptools.org/php_mapscript/
- 9 Perl.com: <http://www.perl.com/>
- 10 UNBC DataShare - Open Source Internet Geo-mapping Framework: <http://datashare.gis.unbc.ca/openigf/>
- 11 独立行政法人情報処理推進機構のフォント (IPA フォント): <http://www.grass-japan.org/FOSS4G/readme-grass-i18n-ipafonts.eucjp.htm>
- 12 Firefox - Rediscover the web: <http://www.mozilla.org/products/firefox/>
- 13 株式会社オークニー: <http://www.orkney.co.jp/>
- 14 株式会社オークニーデータ製品 GIS データパック: http://www.orkney.co.jp/data/datapack_maps.html
- 15 GSI HOME PAGE: <http://www.gsi.go.jp/>
- 16 日本海岸線ポリゴンデータ: http://www.max.hi-ho.ne.jp/scream/coast_pgn.html
- 17 shp2pgsqlのエセ国際化: <http://oi.nu/shp2pgsql/>
- 18 Open Geospatial Consortium, Inc. (OGC): <http://www.opengeospatial.org/>
- 19 郵便番号・市外局番・市町村合併検索「ゆうすけ」: <http://yuusuke.info/>
- 20 全国JIS市区町村名データのダウンロード: <http://yuusuke.info/html/downjis.htm>
- 21 岩手県 合併・名称変更自治体一覧: <http://yuusuke.info/03/03-j.htm>
- 22 さいたま市 指定都市後の郵便番号: <http://yuusuke.info/11/N11100.htm>
- 23 江別経済ネットワークのHPへようこそ: <http://www13.ocn.ne.jp/~eknet/>
- 24 第1号プロジェクト 江別ブランド事典: <http://www13.ocn.ne.jp/~eknet/project1.html>