

UGC を用いた地域資源情報発信システムの提案

佐々木 研弥[†]

窪田 諭[‡]

市川 尚[‡]

阿部 昭博[‡]

[†]岩手県立大学大学院ソフトウェア情報学研究所

[‡]岩手県立大学ソフトウェア情報学部

本研究では、地方自治体の観光促進や地域活性化を目的として、UGC(User-Generated Content)を用いて地域資源情報を収集・発信するシステムを提案する。地域資源情報の収集において、ユーザからの投稿情報だけでは十分な情報が収集できないと考えられるため、ブログから情報抽出を行う。ブログからの情報抽出には地域資源辞書を用い、地域資源について書かれている記事を地域資源情報として抽出する。また、投稿情報とブログから抽出した情報を地図上に表示する機能を開発し、自治体職員を対象に評価を行う。そして、本システムの運用形態についての提案と議論を行い、その結果について考察する。

Suggestion on Regional Resource Information System Using User-Generated Content

Kenya Sasaki[†]

Satoshi Kubota[‡]

Hisashi Ichikawa[‡]

Akihiro Abe[‡]

[†]Graduate School of Software and Information Science, Iwate Prefectural University

[‡]Faculty of Software and Information Science, Iwate Prefectural University

In this study, a system is proposed to collect and send the regional resource information using UGC(User-Generated Content) for communal sightseeing promotion and local revitalization. This system extracts the information from blog, because it is difficult to collect enough information by the user's post information. The system has three functions; (1) to extract the information from the regional blog, (2) to post the information from sightseer and community resident, (3) to display the information using Google Maps. The prototype system is evaluated by local official, and its design conception configured. And, this paper considers the operational model of this system.

1. はじめに

近年、着地型観光と呼ばれる観光形態が注目されている[1]。着地型観光では、着地(観光地)側が観光客にツアーを企画・提案するため、地域の魅力を最大限アピールすることができる。しかし、アピールすべき地域の魅力、つまり地域資源の発信情報量が不足しているのが現状である。地域資源とは、地域に存在する自然資源や人的資源、文化資源など、地域活性化の素材となるものの総称である。地域資源情報の発信を強化することで、観光客数の増加や地域住民による地域資源の保護・運用活動の促進など、地域活性化に繋がる効果が期待できる。

ただし、着地の旅行会社や観光協会などの職員だけで、地域に存在する全ての地域資源の情報を収集することは困難である。そこで、UGC(User-Generated Content: ユーザ生成コンテンツ)を用いた地域資源情報の収集が必要になると考えられる。UGC は、一般ユーザがブログや SNS, Wiki によって生成・公開した情報コンテンツの総称である。特に、国内のブログ総数は 1690 万と非常に多く、月に一回以上更新されて

いるアクティブなブログだけでも 300 万にも及んでいる[2]。

地域の情報を収集する研究として、大槻らの研究[3]と石野らの研究[4]がある。大槻らの研究では、特定地域に関する情報を提供している地域サイトから情報を収集し、地域情報ディレクトリを自動編集するシステムを提案している。しかし、この研究でいう地域サイトとは都道府県や市町村といった自治体の公式 HP のことを指しており、UGC の利用を想定していない。

また、石野らの研究ではブロガーが投稿した日記形式のブログ記事から、観光情報を抽出するシステムを提案している。しかし、観光情報の抽出を「地域名+土産物」という単語の組み合わせで行っているため、土産物について記述されていないブログ記事は抽出されず、具体的な位置情報(緯度・経度)を得ることは難しい。

そこで、本研究では岩手県遠野市をフィールドとし、地域資源情報の発信を目的として、UGCを用いた地域資源情報を抽出・発信するシステムを提案する。また、本システムを運用する際の留意点などについて検討し考察する。

2. 対象フィールドの分析

2.1 フィールド特性

岩手県遠野市は、2010年6月に発刊100周年を迎えた柳田國男の著書「遠野物語」で知られる観光地であり、遠野物語に関連した地域資源が多く存在している。その他にも、遠野遺産認定制度[5]によって、地域資源の掘起しや保護活動が積極的に行われている。遠野遺産には、遠野の文化を象徴する神社などの有形文化遺産や、しし踊りなどの無形文化遺産が99件登録されている(2010年11月現在)。

2.2 ヒアリング調査

地方自治体の情報発信についての現状を知るために、2009年6月に遠野市文化政策部とNPO法人遠野物語研究所を対象に、ヒアリング調査を行った。

その結果、遠野物語100周年記念HPや観光協会HPなどのWebページ上で発信している情報量の不足、特に若い世代へのアピール不足を感じていることを確認した。また、遠野物語関連スポットや遠野遺産など、数ある地域資源情報をコンテンツとして整備するための予算も人員も不足していることが分かった。

そこで、地域住民や観光客に情報を提供してもらう口コミ型のシステムを提案したところ、口コミ型での地域資源情報の収集は有効と考えられるとの意見を得た。ただし、情報提供者はPCや携帯電話などの操作に不慣れな中高齢者が中心になると想定されるため、継続的に情報提供を行えない可能性が高いことが分かった。

3. システム設計

3.1 設計方針

本システムは、地域資源情報の収集・発信を通しての観光客増加と、地域資源の掘起しを目的としている。対象フィールドの分析より、地域資源情報の収集はユーザからの情報提供による口コミ型で行うことが妥当と考えられるが、ユーザから投稿される情報だけでは十分な地域資源情報を収集できない。そこで、地域密着型のブログ(以下、地域ブログ)に蓄積されたブログ記事を、地域資源情報の収集に利用する。地域ブログには、地域のイベントや建物・風景などの地域の情報が写真つきで蓄積されているため、口コミ情報だけでは不足する地域資源情報を十分に補うことができると考えたためである。

したがって、対象ユーザとしては、主にインターネットでの情報収集を行う若い世代の観光客を想定する。また、地域資源の掘起しを通して地域を見つめなおす効果を期待し、地域住民も対象ユーザと想定する。

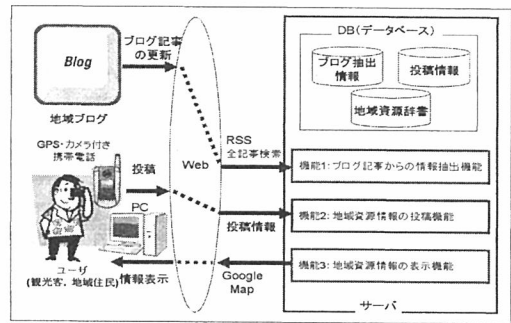


図 1:システム概念図

3.2 システム概要

本システムは、3つの機能で構成される(図 1)。各機能の概要を述べる。

(1) ブログ記事からの情報抽出機能

地域資源について書かれているブログ記事を探し出し、ブログ記事のタイトルや URL、タイムスタンプなどの情報をブログ抽出情報 DB(データベース)に登録する機能である。地域資源名には、地域特有の地名や名詞が使われている場合が多い。正確な情報抽出を行うために、地域資源の情報を登録した地域資源辞書を用いて検索を行う。地域資源辞書に登録される情報は、地域資源名・別表記(複数)・読み(ひらがな表記)・緯度・経度などである。地域資源の中には複数の表記があるため、別表記を登録する(例:しし踊り-獅子踊り-鹿踊り)。

地域資源辞書はシステム管理者が管理するが、ユーザが新たな地域資源を地域資源辞書に登録することもできる。

(2) 地域資源情報の投稿機能

ユーザが携帯電話で撮影した位置情報付きの写真と紹介文を投稿する機能である。投稿は以下の様式に従い、携帯電話のメールで投稿する。

件名: 地域資源名を含むタイトル

添付ファイル: GPSによる位置情報付き写真

本文: 地域資源の解説・紹介・感想など

投稿された情報は投稿情報DBに保存される。また、投稿された情報が地域資源辞書に未登録の地域資源に対しての投稿である場合、新たな地域資源として辞書登録され、ブログ記事からの情報抽出が行われる。

(3) 地域資源情報の表示機能

Google Map上に地域資源の位置をマーカーとして表示し、ブログ情報と投稿情報の2種類の情報を表示する機能である。ブログ情報はブログ記事のページを新規ウィンドウに表示し、投稿情報は投稿情報表示用のページで表示する。

4. プロトタイプ開発と評価

4.1 開発概要

プロトタイプ開発にはLAMP環境を用い、Google Mapの表示部分にはHTMLとJava Scriptを用いた。プロトタイプで用いた地域資源辞書には、所在地が明確な遠野遺産88件を登録した。プロトタイプ開発に先立って実施した基礎実験[6]の結果から、ブログ記事を地図上にマッピングするために既存API[7]を用いるだけでは、地域資源情報を十分に収集できないと判断した。そこで、地域資源辞書を用いて地域資源情報を抽出するプログラムを構築し、ブログ記事から情報を抽出した。

地域資源辞書をもとに、1つのブログ(記事総数448:2009年10月28日時点)に対して地域資源情報の抽出を行った。抽出対象のブログは、本システムの対象とする遠野市観光協会HPからリンクされている遠野の地域ブログ16件のうちの1件である。448エントリーに対して地域資源名を検索し、地域資源名が検出されたブログ記事の情報をブログ抽出情報DBに登録した。その結果、448エントリー中約30%の133エントリーを地域資源情報として抽出した。1つのエントリーから複数の地域資源情報が抽出される場合もあるため、抽出データ数は192件となった。抽出された133エントリー全てに対して、地域資源名が記述されていることを確認している。また、地域資源88件のうち44件が抽出され、44件全てが地図上にマーカーとして表示されることを確認した。

4.2 一次評価

システムの設計方針の妥当性を確認するために、2010年7月15日に行われたいわて観光情報学研究会第5回例会において一次評価を行った。参加者は、遠野市文化政策部、遠野市観光協会などである。そこでは、本システムの3つの機能について説明し、情報抽出の手法と抽出結果の表示画面に対しての評価を行った。

その結果、遠野市にあまり訪れない20~30代の観光客を呼び込むためには良いシステムだという意見を得られたため、システムの設計方針の妥当性を確認した。また、一次評価やそのほかの研究発表で得た意見から、以下のような知見を得た。

(1) 地域資源情報の関係性抽出

ブログ記事に対して形態素解析ツールを用いた文章解析を行い、地域資源間の関係性を抽出する。同じ記事中に登場した複数の地域資源には何らかの関連性があるものと判断する。この関連性を、ブログ記事中で話題の中心となっている地域資源名を特定する手がかりにすること

や、情報表示時に関連のある地域資源へのリンクを表示することで情報抽出・表示に役立てることを考えている。

(2) 情報表示の改善

Google Map上に表示される情報が多くなると、ユーザが必要とする情報を入手しにくくなる。そこで、地域資源を遠野遺産認定制度に準じ「有形遺産」、「無形遺産」、「自然遺産」、「複合遺産」、「その他」に分類し表示する。また、アクセス数の多い情報を有用な情報と仮定し、表示の優先度を決定する。

(3) 運用コスト

抽出されたブログ記事や投稿情報の内容を確認する作業が定常的に発生する。情報量が増えると、その負担は非常に大きくなると予想されるため、作業負担を軽減する仕組みや運用面での工夫について検討する必要がある。

4.3 プロトタイプ改善

一次評価の結果をもとに、プロトタイプシステムを改善した。以下に、一次評価からの改善点について述べる。

(1) 地域資源辞書の更新

遠野遺産の情報更新に伴い、遠野遺産99件全てを地域資源辞書に登録した。地域資源辞書の項目は、後述するMeCab[8]辞書の様式に従う。また、追加エントリーとして地域資源の分類・緯度・経度・関係性(ID)の4項目を新たに登録している(表1)。4.1で行った情報抽出でも十分な数の抽出数を得られたため、地域資源辞書の項目は妥当であると考えられる。

ブログ記事からの地域資源情報と関係性の抽出は、地域資源名・別表記(複数)・原形(正式名称)に登録されている単語を用いて行う。原形(正式名称)は、遠野遺産に登録されている地域資源の名称を指す。正式名称が複雑である場合、情報抽出が困難なため、簡略化した地域資源名を登録する。別表記には、誤表記を含めた別称などを登録する。別表記の項目は、必要に応じて追加可能である。

(2) 複数のブログからの情報抽出

一次評価では、1件のブログ(448エントリー)を対象に、地域資源情報の抽出を行ったが、改善後はブロガーの興味の対象や記事内容の傾向が近く、更新頻度も高い3件のブログを対象とした。2009年10月から2010年9月までの1年間に投稿された692エントリーに対して情報抽出を行った(表2)。ブログAは、一次評価時に抽出対象とした地域ブログである。抽出記事数は、地域資源情報が抽出されたブログ記事の数を指し、抽出データ数は1件のブログ記事から複数

表 1：地域資源辞書の内容(一部抜粋)

ID	地域資源名	別表記 1	別表記 2	原形 (正式名称)	読み (ひらがな)	読み (カタカナ)	分類	緯度	経度	関係性 (ID)
22	カッパ淵	かっぱ淵	河童淵	カッパ淵～蓮池川水域	かっぱぶちとはせきがわすいいき	カッパブチトハセキガスイイキ	有形遺産	39.3542	141.5706	36
27	青笹しし踊り	青笹獅子踊り	青笹鹿踊り	青笹しし踊り	あおざししおどり	アオザサシシオドリ	無形遺産	39.31159	141.569	26,48, 59,70
36	伝承園	*	*	伝承園とその周辺	でんしょうえんとそのしゅうへん	デンショウエントソノシュウヘン	複合遺産	39.3573	141.5692	22,42
59	青笹八幡宮	*	*	青笹八幡宮	あおざさはちまんぐう	アオザサハチマンゲウ	有形遺産	39.31958	141.5684	26,27, 48,85

表 2：ブログ抽出結果

	抽出記事数 (抽出データ数)	記事総数
ブログ A	44(61)	149
ブログ B	14(18)	258
ブログ B	15(23)	285
合計	73(102)	692

の地域資源情報が抽出された場合の重複も含めたデータ数を指している。

(3) 地域資源間の関係性抽出

形態素解析ツール MeCab を用いてブログ記事と投稿情報の文章を形態素に分割し、文章中に含まれる地域資源間の関係性を抽出した。MeCab の標準辞書では、地域資源名のような固有名詞を単語として認識できないため、地域資源辞書を MeCab の標準辞書に反映させている。

1 件のブログ記事または投稿情報の本文中に記述されている地域資源名同士の距離を測定するプログラムを構築し、近い距離に記述されている地域資源同士には関係性があると判断し、関係性抽出の結果を地域資源辞書に反映させている。今回は、複数の地域資源が記述されているブログ記事が 18 件と少ないため、文章中に複数の地域資源名が記述されている場合は距離に関係なく関係性を抽出した。

関係性抽出の結果、表 1 に示す「青笹しし踊り (ID27)」は、「綾織しし踊り (ID26)」「早池峰しし踊り (ID48)」「青笹八幡宮 (ID59)」「細越獅子踊り (ID70)」の 4 つの地域資源との関係性が抽出されている。同じ無形遺産である他地域のしし踊りとの関係性が抽出されているほか、有形遺産である青笹八幡宮との関係性も抽出された。このことから、青笹しし踊りが青笹八幡宮に奉納されるなどの関係性があると推測できる。

(4) 情報表示画面の改善

地域資源の分類ごとにマーカーの色を変更表示し、分類ごとに絞り込めるようチェックボックスを配置した(図 2)。

また、関係性抽出の結果を反映し、マーカーをクリック時に表示される情報ウィンドウ内と、投稿情報表示画面に、その地域資源との関係性が抽出された他の地域資源名が関連スポットとしてリンク表示される(図 3)。

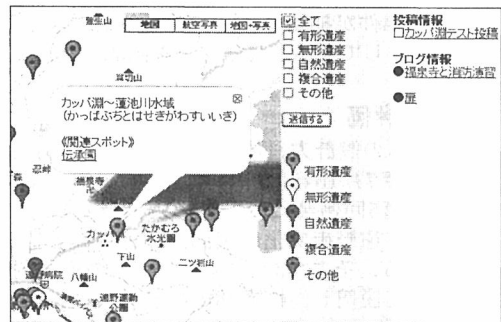


図 2：地域資源の情報表示画面

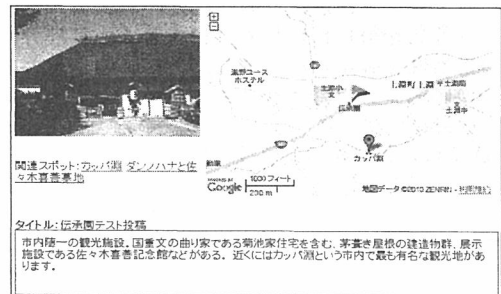


図 3：投稿情報表示画面

4.4 運用形態の提案

本システムは、特定地域での利用に特化しているため、地域特性を考慮した運用形態が必要である。そこで、2.2 のヒアリング調査から得たフィールド特性の分析結果をもとに、本システムの運用形態を提案する。

本システムを地方自治体が主体となって運用・保守することは、作業量の面から考えて困難であるため、本システムの運用主体は地域活性化や観光促進の活動を行っている NPO 法人を想定する。また、地域資源情報の管理負担を軽減するために、地域資源情報の投稿機能は NPO 法人のみが用いるものとし、当面は観光客や地域住民からの情報投稿を制限する。地域資源と密接に関わる遠野市と遠野市観光協会は運用協力する体制とし、NPO 法人と緩やかに連携する。また、ブログ記事からの情報抽出機能は遠野市観光協会 HP からリンクされる遠野関連ブログのみを対象とする。

システム運用時に発生する運用主体の作業としては以下があげられる。

- 地域資源辞書の保守

地域資源辞書への追加・変更・削除を行う。追加と変更は、年に 1 度行われる遠野遺産の追加認定に合わせて行う。地域資源辞書への追加は投稿機能からも行えるため、直接地域資源辞書への追加を行う頻度は高くはないと考える。削除に関しても、不適切な地域資源辞書への追加があった場合に行うため頻度は低い。
- 地域資源情報の管理

運用主体や運用協力者の判断で、不適切な内容が含まれるブログ情報、または投稿情報を削除する。
- 地域資源情報の投稿

お薦めの場所や店などを地域資源として投稿

し、地域資源の掘起しを行う。情報の投稿や、ブログ記事からの情報抽出が行われていない情報を優先的に投稿する。

4.5 二次評価

2010 年 10 月 28 日に、遠野市文化政策部の職員 2 名を対象として、改善したプロトタイプの評価とシステムの運用形態についての提案と議論を行った。運用形態は、本システムが実運用可能な段階にあるという仮定のもとで設定した。二次評価で得た意見を表 3 に示す。表では、改善したプロトタイプの評価と、提案した運用形態についての意見をそれぞれ整理した。

5. 考察

二次評価で得た意見をもとに、システムの今後の課題と運用形態について考察する。

5.1 機能面の考察

- システム全体

現状では、ユーザは地域資源辞書に登録されている地域資源の名前や件数を知ることができず、目的の地域資源の情報を閲覧したい場合にも地図から探す必要があるなど、ユーザの用途に合わせた情報表示に課題がある。システム運用時には、中高齢者の利用が想定されるため、地図画面の操作や情報の投稿方法、得られる情報について説明する必要がある。また、地域資源辞書に登録されている地域資源名を一覧表示し、ユーザが欲しい情報を地域資源名から探す画面が必要である。
- 情報表示画面

現在採用している遠野遺産認定制度に準拠し

表 3：二次評価で得た意見

機能面	【システム全体】	【関係性抽出】
	【情報表示画面】	【ブログからの情報抽出】
	【Google などの検索サービスとの差別化】	
運用面	【運用主体・運用協力】	【情報投稿と抽出対象ブログの制限】

た分類は、大まかな分類であるため、ユーザの利用目的次第では不要な情報が混じる可能性が高い。例えば、神社に興味があるユーザにとって、石碑や水車小屋などの情報まで表示されてしまう「有形遺産」という分類では、欲しい情報を手に入れにくい。そのため、ユーザの興味に対応できる詳細な分類が必要になる。

● ブログからの情報抽出

本システムでは、NG ワード辞書を用意しており、NG ワード辞書に登録された単語がブログ記事または投稿情報から検出された場合、その情報を除外する処理を行っている。NG ワード辞書には、「葬式」「葬儀」など、人の死に関わる単語が登録されている。しかし、遠野物語には遠野市の暗い側面も描かれており、それらの単語を除外すると遠野物語に関する有益な情報を漏らしてしまう可能性がある。例えば、遠野物語で語られる「ダンノハナ」という場所は、現在も使われている共同墓地である。遠野物語には、その他にも老人を捨てた山の話、馬を殺した話、祖母を殺した話などがあり、遠野市のフィールド特性を考慮すると、これらの情報は可能な限り漏らさない NG ワード辞書の設定が必要である。

● 関係性抽出

地域資源間の関係抽出を行うことで、観光客が気づかない関係性、地元住民でも気づかない意外な関係性を抽出できる可能性がある。関係性抽出の質を向上するため、名詞(地域資源名)のみでなく動詞や形容詞を用いた文章内容の分析を行う。

● Google などの検索サービスとの差別化

Google 検索でも地域資源の情報と位置を知ることができるが、本システムでは検索した地域資源の周辺に存在する他の地域資源についても知ることができ、検索サイトでは出てこない地域資源間の意外な関係性について知ることができる。

5.2 運用面の考察

● 運用主体・運用協力

システムの運用・保守の作業量から考えて、自治体職員による運用は難しいため、NPO 法人などの外部組織に依頼しての運用が妥当である。

● 情報投稿と抽出対象ブログの制限

本研究で提案した運用形態では、管理負担を軽減するため、情報投稿者と地域資源情報の抽出対象のブログを制限した。しかし、その結果

本システムで収集可能な情報は地域住民を中心とした、地域に詳しい人たちの情報に限られてしまう。地域の魅力を最も知っているのは地域住民かもしれないが、地域の外の人たちにしか分からない地域の魅力があるはずである。そこで、情報投稿者と対象ブログの制限を段階的に外していくことを考えている。

6. おわりに

本研究では、岩手県遠野市を対象フィールドとして、UGC を用いた特定地域の地域資源情報を収集・発信する手法と、運用形態の提案を行った。二度のプロトタイプシステムの評価から、設計方針の妥当性を確認し、本システムの期待できる効果と課題を明らかにした。

今後の課題としては、動詞や形容詞に注目した文章内容の分析を行った情報抽出・関係性抽出を行い、抽出の精度を向上することが挙げられる。また、情報投稿者と対象ブログの制限を段階的に外すことを考慮し、管理負担を軽減する仕組みを構築する必要がある。

謝辞

本研究の一部は科研費 20500230 の助成を受けています。また、本研究を遂行するにあたり、ご協力いただいた遠野市文化政策部、NPO 法人遠野物語研究所に感謝致します。

参考文献

- [1] 香川真編・日本国際観光学会監修：観光学大事典，木楽舎，2007.
- [2] 総務省 情報通信政策研究所：ブログの実態に関する調査研究，2009.
- [3] 大槻洋輔，佐藤理史：地域情報ウェブディレクトリの自動編集，情報処理学会論文誌，Vol.42，No.9，pp.2310-2318，2001.
- [4] 石野亜耶，難波英嗣，田熊遥，尾崎貴紘，小林大祐，竹沢寿幸：旅行ブログからの観光情報の自動抽出，電子情報通信学会第 15 回 Web インテリジェンスとインタラクション研究会，pp.19-23，2009.
- [5] 遠野遺産認定制度：遠野市公式 HP，<http://www.city.tono.iwate.jp/>
- [6] 佐々木研弥，窪田諭，市川尚，阿部昭博：UGC を用いた地域資源情報発信システムの検討，情報処理学会第 72 回全国大会，4ZL-2，2010.
- [7] LocoStricker：<http://locosticker.jp/>
- [8] MeCab：<http://mecab.sourceforge.net/>