

地域情報の構造化支援とコミュニケーション

熊谷 潤[†] 金杉 洋[†] 柴崎 亮介[†] 山口 琢[‡]

[†]東京大学 空間情報科学研究センター 〒277-8568 千葉県柏市柏の葉 5-1-5

[‡]株式会社ジャストシステム 〒107-8640 東京都港区北青山 1-2-3

E-mail: [†]{kumajun,yok,shiba}@csis.u-tokyo.ac.jp, [‡]taku.yamaguchi@justsystems.com

あらまし 地方自治体や旅行会社、ブログなどに散在する地域データや観光データ、旅行データなどから価値ある情報を広く収集、抽出、構造化し、ユーザ間で簡単に利用できるようにするための研究を行う。具体的には、旅行などをはじめ地域を巡る行動経路を共有可能な応用アプリケーションと、地域データ構造化のためのテキスト入力の簡易性とXMLの構造化拡張性を兼ね備えた新しい入力支援手法の開発を行った。

キーワード 地域情報、観光情報、構造化支援

Study on Input Method for Structuring of Regional Information and Communication

Jun KUMAGAI[†] Hiroshi KANASUGI[†] Ryosuke SHIBASAKI[†] Taku YAMAGUCHI[‡]

[†]Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi Chiba, 277-8568 Japan

[‡]JUSTSYSTEMS CORPORATION 1-2-3 Kitaaooyama, Minato-ku, Tokyo, 107-8640 Japan

E-mail: [†]{kumajun,yok,shiba}@csis.u-tokyo.ac.jp, [‡]taku.yamaguchi@justsystems.com

Abstract Now days, CGM like BLOG, local area SNS on is spreading out in all generations, and the amount of a data which contained human flow data like diary or travel is increasing. Enterprises or the government tried to extract valuable information for them from these data. We study on how to structure easy input system for everyone, and the possibility of reuse and distribution of the regional information and this human flow information. As for the input method, we developed a new input system which contains both usual way of text input and structuring extensibility of XML. Also through a web application that handles the regional information and the human flow information, we study on the possibility of reuse and distribution of these informations.

Keyword Regional Information, Travel Information, Input Method, Structuring

1. はじめに

私たちが街を散策したり旅行に出かけたりする際、現在テレビや雑誌やネットなど様々なメディアから情報を得ることができる。また旅行会社サイトのツアーデータ、地方自治体サイトの観光データなどをはじめ、様々なサイトで地域に関するデータが提供されている。近年では、ブログや地域 SNS、口コミサイトなどが普及し、様々な人の旅行記やおすすめの観光スポットやレストランについても情報を得ることができる。

しかし、こういった地域に関するデータは、主にテキストや画像、PDF 形式で提供されているのが現状である。また実際の位置情報との関連付けがされていないことが多い。私たちは様々なサイトを巡り歩き、記事を見つけそこから必要な情報を自分で切り出したり、地図で実際の場所を調べたりしなければならない。

こうした状況の中、企業による消費者の行動マーケ

ティング、興味の対象を把握した広告戦略、ユーザ毎のパーソナライズサービス、また自治体による交通調査や観光戦略などをはじめ、様々な分野で大量のデータの中から必要な行動情報を拾い出し活用したいというニーズも高まってきている。

さらにブログや WEB サイトなどのテキストデータから特定の情報を抽出しようとする研究がなされている。ブログや文書データから場所情報を抽出し実空間との関連付けを行う研究(相良・有川,2004) (佐藤・西山・砂山,2007)がなされている。さらに経路情報を抽出するという研究(群ほか,2006)がなされはじめており、実世界の人間の行動情報を文章データから抽出し活かそうとする傾向も高まってきている。

しかしながら、これらの研究は主にテキストマイニングによる手法であり、項目別の抽出手法の必要性や抽出精度の限界などの課題があげられる。

一方で XML によって、データをタグで囲み構造化する試みがなされている。XML の入力手法を見てみると、XML の直書きエディタや項目入力などがあげられる。しかし、一般人にとって XML を手書きで記入するのは、かなり困難なことである。また移動情報を扱った既存研究の入力手法については、交通行動調査票の手書きのものから、WEBTFP(大藤ほか,2005)や WEB ダイアリーのような WEB 項目入力を使用したものがある。しかし、ユーザにとって数ある項目を入力していくのは面白みがなく面倒であると考えられる。

2. 研究目的

本研究では、散在する地域データや観光データ、旅行データなどから価値ある情報を広く収集、ユーザ間で簡単に利用できる形で抽出構造化し、移動情報の流通コミュニケーションを促進させるための研究を行う。

具体的には、旅行などをはじめ行動データを共有可能な応用アプリケーションと、それを実現するためのテキスト入力の簡易性と XML の構造化拡張性を兼ね備えた新しい入力支援手法の開発を行った。

3. 地域情報を効率よく再利用するために

様々なサイトの地域データや観光データ、ブログデータを収集し、価値ある情報としてユーザ間で再利用するには、まず決まった形式でそれらのデータを構造化する必要がある。

3.1. 移動データの構造化

移動データを構造化する上で、場所に関する情報は必須であるといえる。次に、移動情報を現実の時間軸に即したものとするため、日時情報が必要である。最後に、抽出するテキストデータについて、どこまでの範囲が該当の場所や日時に関する記述なのかの判定や場所情報と日時情報の関連付けを行うためにイベント区切りの情報が必要である。本研究では、場所、日時(開始日時、終了日時)、イベント区切り。この3つの構造化項目を元にテキストデータを構造化する入力支援手法について研究する。

3.2. 入力支援手法

入力支援手法に関して、文章作成において必ず使用されているカナ漢字変換機能に着目した。また外部から場所や日時に関する情報を取り込むために、ユーザ辞書として入力支援システムと外部サーバとの通信による連携に着目した。本研究では、この2つの着目点をもとに、入力支援システムを開発した。

3.2.1. フォーマット

入力支援システムで入力されるフォーマットは、microformats を採用し拡張した。

・場所タグ : `場所キーワード`

・開始時間タグ : `<abbr class="dtstart">日時キーワード</abbr>`

・イベントタグ : `<div class="vevent">イベント内容</div>`

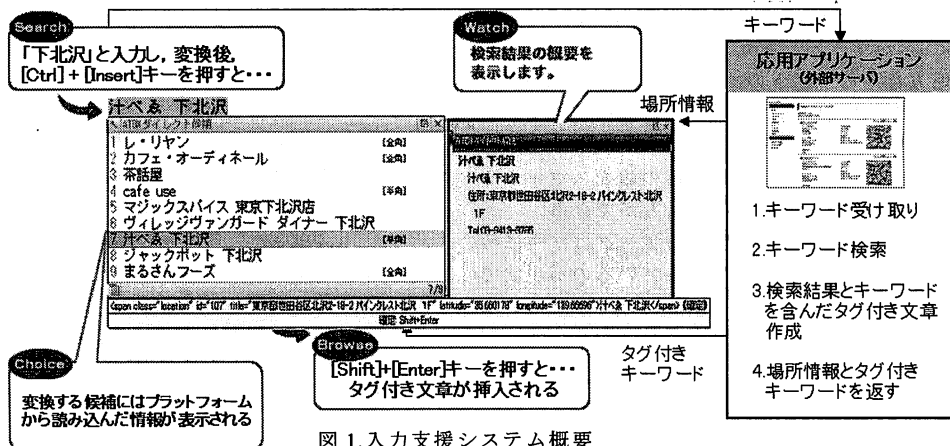
3.2.2. 変換候補値

タグ内に挿入されるキーワードに関する関連情報に関しては、外部サーバとの通信により、関連キーワードについて検索された結果情報を取得し、タグ内に埋め込む。

3.2.3. 場所情報入力支援

ここでは、「下北沢」と入力した場合を例に、入力支援システムの挙動について説明する。

1. 「下北沢」と入力を受け取った入力支援システムは、外部サーバのアプリケーションに API を通じて「下北沢」で検索を行う。
2. アプリケーションは、「下北沢」に関するショップや施設などの場所に関する検索結果を入力支援システム側に API 配信を行う。
3. 入力支援システムは、アプリケーションから受け取った XML の結果を解析し、場所タグ `` 内に返値を挿入し、入力候補値として表示する。
4. ユーザは「下北沢」に関する返値候補から自分の入力したい候補値を選択入力する。



3.2.4. 日時情報入力支援

ここでは、「明日の午後 3 時」を例に入力支援システムの挙動を説明する。

1. 「明日の午後 3 時」と入力を受け取った入力支援システムは、「明日」と「午後 3 時」の語句を抽出し、現在日時から「明日」「午後 3 時」を計算し正しい日時の値を求める。
2. 開始日時タグ<abbr class="dtstart" title="">もしくは終了日時タグ<abbr class="dtend" title="">内の title 属性値に返値を挿入し、入力候補値として表示する。
3. ユーザは、返された候補値から自分の入力したい日時（開始か終了）を選択入力する。

3.2.5. イベントタグ入力

イベントタグの入力時の入力支援システムの挙動は以下の通りである。

1. 「イベント」と入力を受け取った入力支援システムは、開始タグ<div class="vevent">および終了タグ</div>を値として返す。
2. ユーザは、イベント開始もしくは終了にあわせ該当のイベントタグを選択入力する。

3.2.6. 入力支援手法による移動データ構造化例

入力支援システムによって、構造化された移動データの文章データの例を以下に示す。

1. 結果表示

今日 7 時ころから、六本木で映画をみてきた。

2. 結果ソース

```
<div class="vevent"><abbr class="dtstart" title="2008-08-16T19:00">今日  
7時ころから、<abbr class="location" id="2" name="TOHO シネ  
マズ六本木" latitude="35.660175" longitude="139.728958" title="東京都  
港区六本木 6-10-2 六本木ヒルズ|ナキ坂コンプレックス内">六本木  
</span>で映画をみてきた </div>
```

3.3. 入力支援手法による移動データ構造化メリット

本入力支援システムにより、XML などについて知らない一般ユーザにでも、文章を書く際にタグ変換し移動データの構造化入力が可能になる。

また文字を入力できる箇所なら、どんなアプリ上でもでも使用可能であり、ブログや文書だけでなく、メールやチャットなどにおいても使用可能である。

テキストマイニングでは困難な「青山」など地名か人名なのか単語だけでは区別できない箇所でも、場所と指定し入力できる。

詳細な位置情報として公開したい場合はタグ付きで入力、公開したくない場合は通常入力をするなど、ユーザの判断で情報の公開制御が可能になる。

さらに地名に限らず、様々な専門データベースと連携することにより、それらの情報をタグ内に挿入し入力できる。

4. 応用アプリケーション「モスケ」

地域情報を構造化入力してもらうためには、ユーザにとってのメリットが必要である。そこで本研究では、ユーザが旅行などのおすすめスケジュールをモデルスケジュールとして投稿できる口コミサイト「モスケ」を開発した。地域での移動情報を構造化単位で投稿、収集、検索、参照などが可能で、地域情報のコミュニケーションを促進するアプリケーションとなっている。

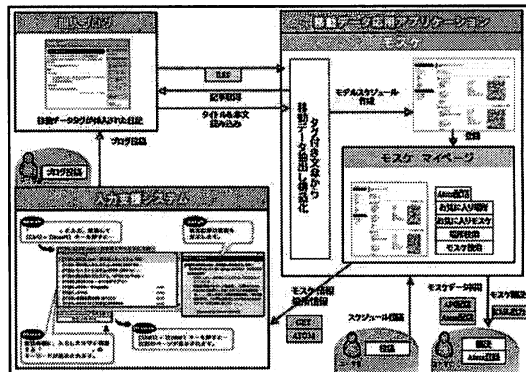


図 2. 「モスケ」システム概要図

4.1. 特徴

「モスケ」では、ブログなどの記事上の移動情報タグを解析し、モデルスケジュールとして変換保存する機能を実装している。「モスケ」上に RSS を指定することにより、様々なブログやサイトからタグ付きの記事を収集解析する。記事内のイベントタグ毎に分割しそれを組み合わせたスケジュール情報として、ブログやニュース収集サイトのように、WEB 上に散在する様々なスケジュール情報を集積することが可能である。

次に特徴的なのが、地名データの API 配信である。「モスケ」に蓄積された地名などの情報は、API を通じて入力支援システムの地名辞書として活用され、緯度経度情報を含んだタグ付き地名として入力できる仕組みになっている。

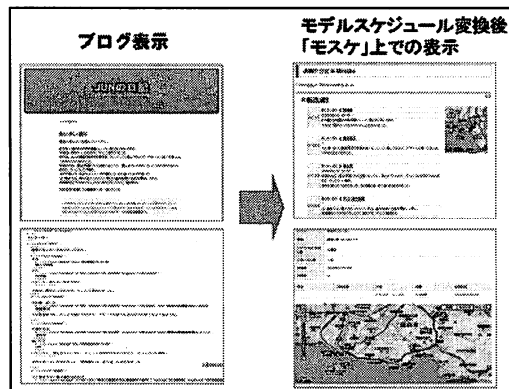


図 3. ブログからの移動情報抽出例

4.2. ユーザによる地域情報のコミュニケーション例

ここでは、「モスケ」を利用した地域情報の活用、コミュニケーションについて、普段から出張日記を綴る出張の達人 T さんと出張先がはじめての J さんの 2 人の人物のやりとりを例に紹介する。

1. 出張の達人 T さん

- ① 出張先のノウハウをブログに入力支援システムを使用して記事を書く。
- ② 「モスケ」にブログの RSS を設定し、ブログを登録。
- ③ 自分の記事の中から、「モスケ」に登録したい記事を登録
- ④ 自分のモデルスケジュールが参照されて、評価を得る。

2. 出張先が初めての J さん

- ④ 出張先のモデルスケジュールがないか出張先の場所で検索を行う。
- ⑤ T さんが登録したモデルスケジュールがとも参考になる！
- ⑥ T さんのモデルスケジュールを元に自分のなりの予定を追加しカスタマイズし、経路などを取得。API を通じてカレンダーにも出力しておく。
- ⑦ 出張当日はそれに従って行動すればいいので便利！
- ⑧ T さんのモデルスケジュールを評価するとともに、今回カスタマイズしたスケジュールを「モスケ」に登録。
- ⑩ 新たに J さんが登録したモデルスケジュールがさらに他のユーザに利用される。

以上のように、ブログを書くついでに文章内の地域データや旅行データを構造化しユーザ間で共有することにより、おすすめスケジュールの追体験やそれを活かした自分なりのスケジュール立案が可能であり、移動情報や地域情報を実生活に活かせるようになる。

4.3. 「モスケ」による地域情報の活用

まずユーザのメリットとして、単なる文章データから抽出し構造化することにより、地域を自分がどのようにまわったのか途中の経路もふくめて地図表示などの詳細な視覚化が可能になる。共有されたモデルスケジュールは場所などの構造単位で検索、参照、転用が可能になり、初めての地域を訪れる際でも、評価の高いモデルスケジュールを参考にすることで直接生活に活かすことができる。

一方で地域の情報を扱う地方自治体や旅行会社にとっては、観光客が地域をどのようにまわるのか、どういう評価なのかといった行動把握が可能になる。また場所だけでなく経路をもとにしたサービスや行動マ

ーケティングやパーソントリップなど動線解析などへの応用も考えられる。さらに蓄積された地名情報やモデルスケジュール例を、API 配信を利用しブログや外部アプリなどへ掲載することにより、地域情報をより多くの人の目に触れ活性化することが可能である。

5. まとめ

本研究では、散在する地域に関するデータや移動データを構造化し価値ある情報として利用するため、慣れ親しんだテキスト入力 of 簡易さと XML の構造化拡張性を兼ね備えた新しい入力支援手法を示した。さらに構造化された移動情報を構造単位での検索や再利用・流通を可能にする応用アプリケーション「モスケ」による活用例やメリットを示した。

応用アプリケーション上に様々な地域情報が蓄積されることにより、ユーザ間で活用されるだけでなく、入力支援システムの地名辞書として再利用することが可能になる。これによりさらに構造化入力が利用しやすいものとなり、構造化された地域情報が流通する。このように本研究の入力支援システム、アプリケーションによって、ユーザが作成した地域情報を簡単に再利用できる仕組みができ、ユーザ間の地域情報のコミュニケーションが促進されると考えられる。

今後、入力支援システムの有効性の検証として、マニュアル時と入力支援システム使用時での構造化入力にかかる所要時間の計測比較、またブログ記事から移動データの抽出精度、ユーザの反応などを実験評価する予定である。

また地域情報の再利用・流通可能性について、特定エリア内の観光データ、ツアーデータ、ブログデータなど様々な移動データを構造化し蓄積することにより地域情報をどのように利用できるか、どのようにユーザ間で利用されるのかを評価検証していく予定である。

文 献

- [1] WEB ダイアリー、
”<http://www.probe-data.jp/demo/diary.html>”(2008/06/27 アクセス)
- [2] 大藤武彦,松場圭一,井上英樹,松村暢彦「WEB を活用したトラベル・フィードバック・プログラムの多様な事業所への適用」土木計画学研究・講演集 CD-ROM Vol.31,2005
- [3] 郡宏志,服部峻,手塚太郎,田島敬史,田中克己”ブログからのビクターの代表的行動経路とコンテキストの抽出”,情報処理学会研究報告,p35-p42,2006
- [4] 相良毅,有川正俊 ”ジオパースによる Web からの空間コンテンツ獲得”電子情報通信学会データ工学ワークショップ,I-11-01,2004
- [5] 佐藤圭太,西原陽子,砂山渡”ブログ記事からのイベント文抽出によるシーンの生成”,人工知能学会,2007