

論文内容の要旨

論文題目 人の移動記述データの構造化入力支援手法と
移動記述情報の流通可能性に関する研究
(Study on Development of Input Method for
Structuring of Human Flow Data and Possibility of
Distribution of Human Flow Information)

氏名 熊谷 潤

1. 背景と目的

私達の周りには、ネット上をはじめデジタルデータが増加し大量のデータから必要な情報を見つけ出すのが困難な状況となっている。原因の1つとして、データ作成時に索引付けや構造化が行われていないことが挙げられる。ここで、コンテンツ単位での構造化やコンテンツ内の情報単位で構造化することにより、情報の検索性や抽出精度の向上が期待される。こうした構造化を行うためには、XMLの直書きやテンプレートへの入力が挙げられる。しかし、直書きが困難な点やテンプレートでは自由に記述できない。またコンテンツ作成過程において、「検索」「抽出」「入力」「構造化」にかかる操作コストが高く、操作毎に課題が独立して存在している。本研究では、1. 「検索」「抽出」「入力」「構造化」を入力時に同時に可能な手法、2. 様々なデータを構造化できるように様々なアプリケーション上で使用可能な要件を満たす構造化入力支援手法を提案する。

コンテンツ作成時において、既存の構造化情報を組み合わせることにより容易に作成できる。しかし、構造化情報を個人や社内、ポータルサイト内の閉じた環境のみで保持しては十分な再活用ができない状況である。そこで本研究では、1. 外部から構造化情報を「収集共有」可能、2. 共有されることによる「付加価値」の提供、3. 構造化入力支援手法に「配信」可能である点を満たす構造化情報活用環境の仕組み作りを提案する。

本研究で提案する構造化入力支援手法と構造化情報活用環境により、データの構造化が簡単になり、構造化情報が増え構造化情報活用環境に収集蓄積される。蓄積された情報を構造化入力支援手法に配信することによって、構造化入力支援手法による入力がより簡単になる。このように

情報活用サイクルが円滑化され、私達の情報の生産性や利便性が向上すると考えられる。

本研究では、以下の観点から対象データを人の移動に関する記述データ（移動記述データ）に設定し実証を行う。1. 観光データや旅行記など人の移動に関するデータが散在している点、2. 位置情報などと関連付けし視覚化などの付加価値を提供できる点、3. 企業や自治体をはじめ、人の移動情報を活用したいというニーズが高い点。

2. 移動記述データの構造化

移動記述データの構造化をする上で、様々な移動記述データを分析した結果、「場所」「日時」「イベント（文章中の場所と日時に関連する記述範囲を指定）」の3つの項目で構造化することを提案する。構造化に必要なフォーマットとして、1. 3つの要素に対応できる、2. 見た目を保持し構造化ができる、3. WEB上で普及している点の3つの条件を満たす microformats の hcalendar を採用した。

3. 移動記述データの構造化入力支援手法

本研究では、テキストマイニング手法の課題である抽出精度の向上、XMLでの構造化入力の困難さの解決を目指し、以下の要件を満たす構造化入力支援手法を提案する。1. 「検索」「抽出」「入力」「構造化」が同時に可能。2. 様々なアプリケーション上で使用可能。3. 「イベント情報」「日時情報」「場所情報」の項目を移動記述情報タグとして入力可能。

文字を入力する際に使用するカナ漢字変換機能、辞書として外部データベースの使用に着目した。これにより、文章入力時に外部データベースから取得した様々な情報（緯度経度情報や正式場所名など）を埋め込んだタグに変換し、移動記述データを文章入力時に構造化が可能になる。構造化入力支援手法を ATOK 上にプラグインとして実装した（図 3.1）。

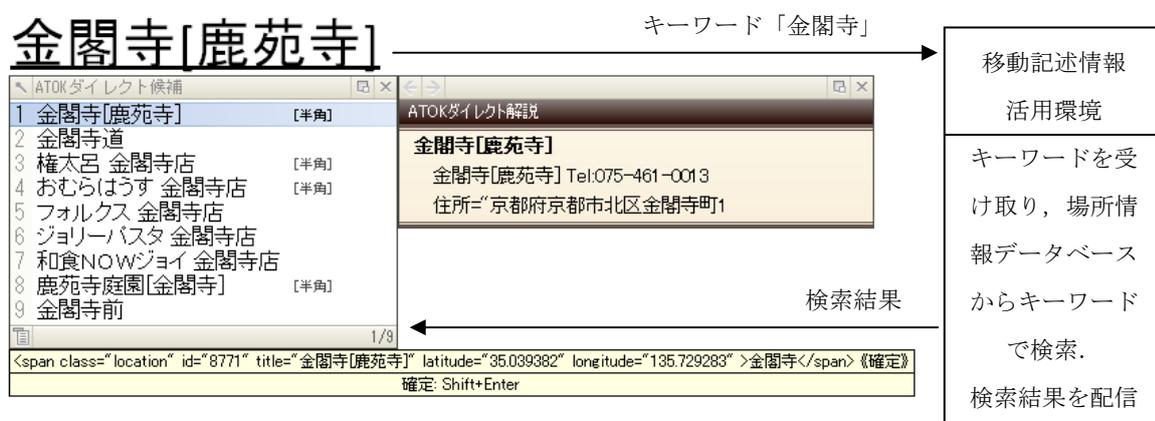


図 3.1 構造化入力支援システム概要



図 3.2 構造化後の文章（通常表示）



図 3.3 構造化後の文章(HTML表示)

構造化入力支援システムの入力作業時間や操作性について評価を行うため、定型文（文字数：296文字，場所タグ数：5，イベントタグ数：10）を使用して、通常入力，構造化入力支援システムによる構造化入力，タグ直書きによる構造化入力の3種類の入力速度比較実験を行った。構造化入力支援システムを使用した場合，通常入力に86秒の追加作業を加えるだけで，タグ直書き入力に比べ約9倍速く文章を構造化することができた（表 3.1）。

表 3.1. タグ抽出結果

	入力速度	通常入力	入力支援システム	タグ直書き	入力支援システム作業時間	タグ直書き作業時間
平均	157 文字	312 秒	398 秒	1050 秒	86 秒	738 秒
経験者	170 文字	285 秒	347 秒	847 秒	71 秒	571 秒
初心者	144 文字	347 秒	448 秒	1251 秒	101 秒	904 秒

操作方法について PC 初心者でも簡単に習得でき, XML の知識がない人にとっても簡単に XML データを入力することができることを示した (図 3.4-3.6). 構造化入力支援システムに対する利用希望度は約 70%と高い結果となった (図 3.7).



図 3.4 操作感

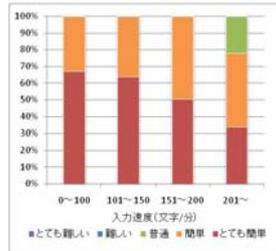


図 3.5 入力速度別

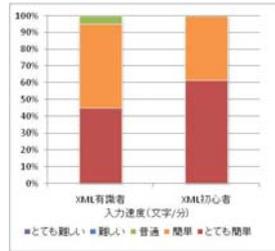


図 3.6 XML 知識レベル別

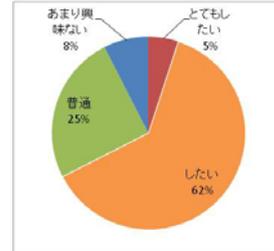


図 3.7 利用希望度

4. 移動記述情報活用環境

移動記述情報を活用するための要件として, 「構造化」「収集」「付加価値」「公開制御」「配信」を定義し, これらを満たす移動記述情報活用環境として, おすすめの移動記述情報をモデルスケジュールと定義し, ユーザ間で共有活用できるロコサイト「モスケ」を作成した.

本研究の移動記述情報の抽出精度の評価を行うため, 40 人の被験者に旅行記を作成・構造化してもらい, 移動記述情報活用環境でタグ解析し抽出結果を確認してもらった. 結果として, テキストマイニング手法と比較して, 適合率, 再現率ともに高い抽出精度を実現できた (表 4.1). 被験者へのインタビューの結果, 利用希望者の割合が過半数を超え, 海外旅行頻度の多い被験者, CGM 利用者の利用希望度が高い結果となった.



図 4.1. 概要図

表 4.1 タグ抽出結果

	適合率	再現率
本研究	100%	94.9%
テキストマイニング	83.2%	71.5%



図 4.2 ブログからの移動記述情報抽出

5. 移動記述情報の活用手法

移動記述情報活用環境上に移動記述情報が共有されることによって得られる参照数や場所間の共起関係情報を活用する手法について述べる. 使用データとして, 被験者の旅行記 40 件, 観光雑誌, 京都市サイト, ポータルサイト, バスツアー, タクシーツアーの各カテゴリから 100 件ずつ

移動記述データを構造化し、合計 440 件の移動記述情報を使用した。

移動記述情報からの参照数を活かす手法として、人気場所として参照数が多い場所を示した(表 5.1)。また、参照数の多い順に移動記述情報活用環境内の空間データベースの結果をソートすることにより、構造化入力支援システムの操作性を向上できる(表 5.3)。

表 5.1 参照数 (全体)

順位	場所名	参照数
1 位	清水寺	61
2 位	金閣寺	49
3 位	天龍寺	45

表 5.2 検索候補結果

検索語	目的語	共有前		共有後	
		順番	時間	順番	時間
京都	京都駅	45 番	17 秒	3 番	7 秒
嵐山	嵐山駅	46 番	20 秒	3 番	6 秒
河原町	河原町駅	486 番	圏外	5 番	9 秒

次に利用者の経路選択についての分析や場所情報推薦を行うため、N-gram によって移動記述情報に含まれる隣り合う場所間の共起関係を求めた(表 5.3)。被験者の旅行記と各メディアを比較するため共起頻度を求めた。参照数の多い人気観光場所を効率的に巡るタクシーやバスツアーの移動記述情報と類似している結果となった(図 5.1, 図 5.2)。

表 5.3 共起関係 (全体)

順位	場所名	数
1 位	永観堂⇔南禅寺	16
2 位	龍安寺⇔仁和寺	13
3 位	金閣寺⇔龍安寺	12
3 位	詩仙堂⇔曼殊院	12

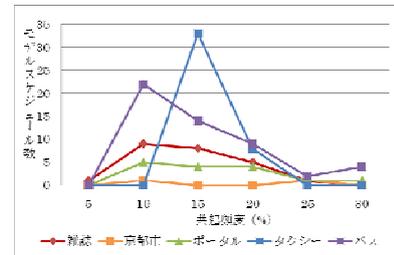
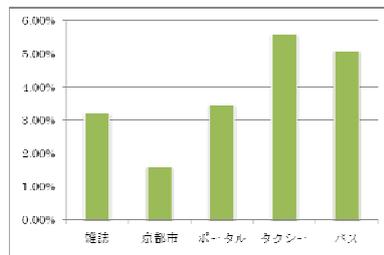


図 5.1 旅行記との類似度

図 5.2 旅行記との類似度

また場所毎に共起関係を集計し、各場所に「どこから訪れる人が多く、どこへ行く人が多いのか」を解析した(図 5.3, 図 5.4)。その結果を利用し、開始場所(例. 清水寺)を指定し共起関係の強い場所を選択していくことによってモデルスケジュール案を示した(図 5.5)。移動記述情報を共有し解析することにより、参照情報以上に、利用者へ有益な情報として還元可能である。

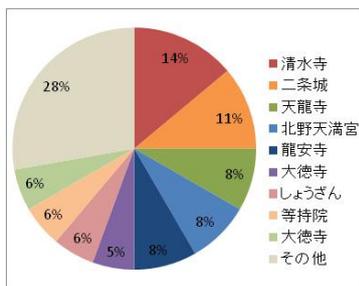


図 5.3 X→清水寺

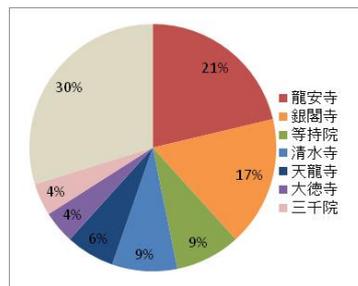


図 5.4 清水寺→X



図 5.5 清水寺始点のモデルスケジュール

6. 結論

本論文では、散在する様々な移動記述データから移動記述情報として生産性や利便性を向上させるべく、構造化入力支援手法による移動記述データの構造化入力支援手法、移動記述情報活用環境による移動記述情報活用の仕組みについて述べた。これにより、移動記述情報の活用サイクルが形成され、誰にでも簡単に移動記述情報を活用できる仕組みが実現できたといえる。

今後の展望として、構造化入力支援手法に関しては入力時の利用者の「時間」「場所」「使用環境」に最適な入力候補値を推薦する手法、構造化情報活用環境に関しては個人履歴やコンテキストに基づいた各個人にとって最適な移動記述情報の推薦手法の研究が有効と考えられる。