

小説や旅行記からの地理情報の自動抽出と編纂

平山 拓実[†] 難波 英嗣[‡] 竹澤 寿幸[‡]

[†]広島市立大学 情報科学部 〒731-3166 広島県広島市安佐南区大塚東 3-4-1

[‡]広島市立大学大学院 情報科学研究科 〒731-3166 広島県広島市安佐南区大塚東 3-4-1

E-mail: ††{hirayama, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp

あらまし 近年、ストーリー性のある情報媒体の舞台場所を訪れる観光客が増加している。このような観光では、多大な経済効果をもたらすとして注目されている。本研究では、書籍の舞台場所を訪れる観光客の支援として、書籍のストーリーと関連のある舞台場所を可視化するシステムを構築する。本システムは、ジオコーディングとテキスト要約から構成する。ジオコーディングでは、書籍における都道府県の出現頻度を用いて曖昧性問題の解消を行っている。また、書籍のストーリーと関連のある舞台場所を抽出するため、テキスト要約を用いる。本システムにより、書籍における舞台場所の確認が容易となり、観光客の増加が望める。

キーワード 可視化, ジオコーディング, 経路抽出, テキスト要約

1. はじめに

本稿では、書籍のストーリーと関連のある舞台場所を提示するシステムを構築する。近年、ドラマやアニメ、書籍といったストーリー性のある情報媒体における舞台場所の観光が増加している。有名な例として、NHK朝ドラの「あまちゃん」や大河ドラマ「龍馬伝」が挙げられる。このような舞台場所を観光する旅行者により、「あまちゃん」の岩手県にて約33億円¹、「龍馬伝」の高知県にて535億円²、長崎市にて191億円³の経済波及効果をもたらしている。このように舞台場所の観光では、多大な経済効果をもたらすとして注目されている。

ドラマやアニメでは、ホームページにストーリーと関連のある舞台場所を可視化した情報が記載されていることがある。そのため、舞台場所の確認が容易にできる。しかし、書籍では可視化された情報がない、という問題点がある。そこで本研究では、書籍のストーリーと関連のある舞台場所を地図上にマッピングすることにより、可視化するシステムを構築する。

地図上にマッピングするには、舞台場所を示す地理情報にジオコーディングする必要がある。しかし、書籍における舞台場所を示す地理情報は、正式な地名ではなく、省略された地名で記載されることが多い。そのため、どの都道府県に属する地名であるか不明瞭となる、曖昧性問題がある。本研究では、このような曖昧性のある地名に対して、書籍における都道府県の出現頻度を用いることにより、最も適切と思われる緯度経度を付与する手法を提案する。

ストーリー性のある情報媒体における舞台場所の観光では、主人公が辿った場所やストーリーと関連のある場所を舞台場所として観光することが多い。本システムにおいて提示する舞台場所は、ストーリーと関連のある場所としている。そのため、テキスト要約を用いて、このような場所のみを抽出する。

ジオコーディングのみを用いたマッピングでは、不要な舞台場所を多く含む提示となる。そのため本研究では、マッピングする際に、上記のジオコーディングとテキスト要約の二つの手法を組み合わせる。これにより、ストーリーと関連のある舞台場所のみを提示できるため、書籍のストーリーと関連のある舞台場所の確認が容易となる。

本論文の構成は以下の通りである。2章では本システムの概要・動作例を示し、3章では関連研究について述べ、4章では書籍データからの地理情報の抽出と編纂について述べる。5章では評価実験について述べ、6章で本論文をまとめる。

2. システムの概要・動作例

本節では、構築するシステムの概要と動作例について説明する。まず、本システムの概要について述べる。書籍のストーリーと関連のある舞台場所にマッピングすることにより、舞台場所の確認を容易とする。また、ピンが表示された舞台場所の出現場面を把握するため、一文を提示するシステムである。

本システムでは、青空文庫から収集した6,261件の書籍データを対象としている。書籍名または地名を入力として受け取り、一致するデータを本システムのデータベースから検索する。次に、ストーリーと関連のある舞台場所を地図上にマッピングする。地図上にマッピングしたものを本システムの出力とする。図1に示す出力例のピンをクリックすると、書籍から抽出した地理情報とそれを含む文、緯度経度といった情報が

¹ 岩手経済研究所, 岩手経済研究 25年9月号

² 日本銀行高知支店, NHK大河ドラマ「龍馬伝」の経済波及効果(2011)

³ 長崎市, 平成22年長崎市観光統計(平成23年3月)

提示される。本研究では、情報媒体を対象とした観光では、経路選択も一つの楽しみであると考えている。そのため、経路は提示しないものとする。



図 1. システム動作例

3. 関連研究

本研究の関連研究として、3.1 節では、経路案内システム、3.2 節では、ジオコーディング、3.3 節では、テキスト要約についての研究をそれぞれ紹介する。

3.1. 経路案内システム

本研究のような、情報媒体における経路を提示する研究に Jannik[1]らの研究がある。Jannik らは、電子化した書籍データから抽出した時間情報と地理情報を組み合わせてイベントを生成し、地図上に表示する手法を提案している。Jannik らの研究における、電子化された書籍データを用いている点や、地理情報を抽出する点で本研究と類似している。しかし、地理情報の抽出やジオコーディングに外部 API を用いている。本研究では、地理情報の抽出に外部 API を用いているが、ジオコーディングには生成した地名辞書を基に行っている点で異なる。

中嶋ら[2]は Twitter から情報を抽出し、経路を生成する手法を提案している。この研究では、Twitter における観光ツイートを「食事」、「景観」、「行動」の3つに分類する。さらに、旅行者のタイムラインから観光ルートを抽出し、旅行者の好みに合わせた観光ルートを推薦している。本研究では、これから書籍を対象とした観光を行う観光客の支援として観光場所を提案することを目的としている。そのため、地理情報の緯度経度に対応する位置にピンを立て、舞台場所だけを示している。

3.2. ジオコーディング

地理情報に対してジオコーディングする既存のシステムに、Geocoding Tools&Utilities⁴がある。このシステムでは、地名を入力とし、緯度経度の出力を行う。しかし、このシステムでは、曖昧性のある地名に対し

て、正しい緯度経度の付与ができない。例えば、入力に山梨県や埼玉県、岩手県に存在する「赤沢」を用いた場合、岩手県に属する赤沢の緯度経度しか付与できない。

このような曖昧性のある地名に対して正しい緯度経度情報を付与する研究では、クラスタリングを用いた河野ら[3]の研究がある。河野らは、Twitter⁵におけるユーザの一連の Tweet から地名を抽出し、Tweet の位置情報を推定する研究を行っている。曖昧性のある地名に対応する候補と、前後の Tweet で抽出した地名の両方を対象にクラスタリングを行う。各クラスタに対して、候補と曖昧性のない地名を用いてスコアを付与する。そして、最も総和の大きいクラスタにある候補の緯度経度を適切な緯度経度として付与している。このようなクラスタリングを用いた手法では、距離を用いている。そのため、書籍データの舞台場所となる都道府県と異なる都道府県の地名に緯度経度情報が付与される可能性がある。本研究では、曖昧性のない地名の都道府県を基に、曖昧性のある地名に緯度経度情報を付与している。

曖昧性のある地名に対してクラスタリング手法を用いない研究に金木ら[4]、平野ら[5]の研究がある。金木らは、地名テーブルとランドマークテーブルの二つからなる地名辞書を提案している。これにより、曖昧性を持つ地名に対して地域間距離とランドマーク距離を用いて特定を行うことができる。ランドマークとは地域を特定する際に、地名と同等の意味があるものである。この研究では、ジオコーディングの精度を向上させるランドマークテーブルの作成を目的としているため、本研究とは目的が異なる。平野らは、曖昧性のある地名をジオコーディングする際、地名の距離と有名度を用いる手法を提案している。店の多い場所を有名な場所として、店の件数を有名度のスコアとしている。ジオコーディングでは、地名の距離を優先的に処理し、候補の中で有名度が突出している候補を優先して処理を行う。しかしこの手法では、店の少ない地域の地名では、正しい緯度経度情報が付与されない問題点がある。本研究では、このような事象の影響がない手法である。

3.3. テキスト要約

本研究では、書籍のストーリーと関連のある舞台場所を提示するシステムの構築をするため、テキスト要約を行う。そのため、本節ではテキスト要約に関連する研究について紹介する。これまでに多くの自動要約手法が提案されている。

相良ら[6]は、テキスト中の多くの段落に現れる話題をメインピック、最低 2 つの段落で現れる話題をサ

⁴ <http://newspat.csis.u-tokyo.ac.jp/geocode/>

⁵ <http://twitter.com/>

ブトピックとし、これらのトピックを用いた要約により、書籍のストーリーを作成する手法を提案している。この研究では、要約文の生成には、各トピックを代表する語を含む文を統合する手法を提案している。

Erkan ら [7] は、テキストをグラフで表現し、PageRank アルゴリズムを適用、テキスト中の各文の重要度を測る手法 LexRank を提案している。重要度の算出には、次数の多いノードを評価するだけでなく、次数の多いノードと隣接しているノードの重要度についても考慮し、対象ノードを評価することができる。この手法では、文間のコサイン類似度に基づいた連結性行列が文のグラフ表現の隣接行列として使われており、その隣接行列の第一固有ベクトルの成分を各ノードの中心性を表すスコアと考える。また、LexRank は多くのテキスト要約タスクでその有効性が確認されている。そのため、本研究のテキスト要約には、この LexRank 手法を用いる。

4. テキストデータから地理情報の抽出と編纂

4.1. システム概要

図 2 にシステムの概要を示す。また、本システムの流れを以下に示す。

- (1) 青空文庫を利用し、人手で書籍データを収集
- (2) 固有表現を抽出し、タグ付けを行う
- (3) 地理情報のジオコーディング
- (4) 3.3 節で述べた LexRank 手法を用いたテキスト要約
- (5) (3) と (4) で生成したデータを組み合わせ、データベースとする
- (6) データベースを参照し、入力 of 書籍と対応するデータを出力する

4.2 節では、固有表現のタグ付け、4.3 節では、ジオコーディング、4.4 節では、テキスト要約について説明する。

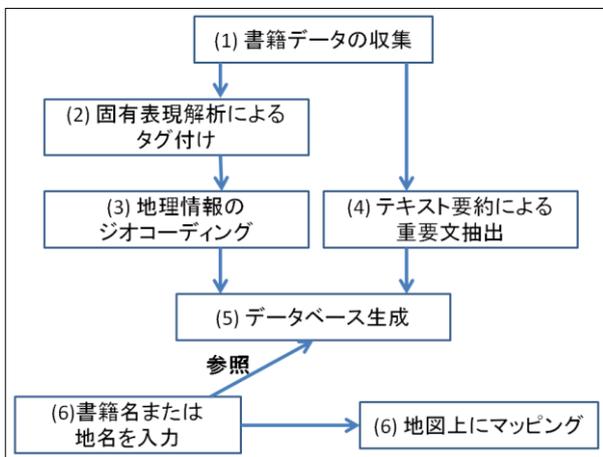


図 2. システム概要

4.2. 固有表現のタグ付け

本節では、固有表現の抽出とタグ付けについて述べる。本研究では、係り受け解析器 CaboCha の固有表現抽出機能を利用し、固有表現を抽出する。CaboCha の固有表現抽出機能では、ORGANIZATION(組織名), PERSON(人名), LOCATION(地名), DATE(日付表現), TIME(時間表現), MONEY(金額表現), PERCENT(割合表現), ARTIFACT(固有物名)の8種類の固有表現を抽出する。本研究では LOCATION または ORGANIZATION と分類された固有表現を地理情報として扱う。また、CaboCha の固有表現抽出機能は、IREX-NE で公開された定義⁶に基づいた固有表現を抽出している。以下に LOCATION と ORGANIZATION の定義を示す。

- LOCATION: 固有の場所を指す名前。大陸, 国名, 地域名, 都市名, 県名, 町名, 道路名, 住所, 駅名, 線路名, モニュメント, 海洋名, 湾, 運河, 川名, 池名, 湖名, 島, 公園, 山, 砂漠の名前などを含む。
- ORGANIZATION: 複数の人間で構成され、共通の目的を持った組織などの名称を指す。株式会社などの会社, 固有の政府組織, 学校, 軍, スポーツチーム, 国際組織, 労働組合, 工場, ホテル, 空港, 病院, 教会や, なんらかの目的を持ったグループなどもその対象が組織としての意味で使われている文脈においては組織名とする。

本研究では、CaboCha の固有表現抽出機能による抽出結果を用いて自動タグ付けを行う。書籍「駅夫日記」にタグ付けを行った結果の一部を図 3 に示す。

この間の工学士というのは<LOCATION>品川</LOCATION>に住んでいた<ORGANIZATION>東京市街鉄道会社</ORGANIZATION>の技師を勤めている<PERSON>蘆鉦次郎</PERSON>という男で、...

図 3. タグ付け例

4.3. 地理情報のジオコーディング

本節では、書籍に出現する地理情報のジオコーディングについて述べる。本研究のジオコーディングには、地理情報と対応する情報を地名辞書から検索する方法を用いている。本研究の地名辞書は、国土地理院⁷から収集した都道府県と市区町村の計 1,957 件、言語資源協会⁸から収集した町丁と施設名の計 117,061 件、

⁶ <http://nlp.cs.nyu.edu/irex/>

⁷ <http://www.gsi.go.jp/>

⁸ <http://www.gsk.or.jp/>

Wikipedia から収集した山や駅名など計 44,930 件のデータで構成される。地名辞書のテーブルは，“名称”，“緯度”，“経度”，“粒度”の 4 項目から構成されている。“粒度”は，“名称”が都道府県の場合は 1，市区町村の場合は 2，これより詳しい場所を指すものには 3 を振り当てている。地名辞書のデータ例を表 1 に示す。“名称”には，地名の正式名，山や駅名の名称が記載されている。地名の正式名は，都道府県，市区町村，町丁の境界部分をスペースで区切る。

表 1.地名辞書に登録されたデータ例

名称	緯度	経度	粒度
北海道	43.06417	141.34694	1
北海道 函館市	41.76861	140.72917	2
北海道 函館市 青柳町	41.75731	140.71667	3
幌尻岳	42.71944	142.68277	3

書籍データにおける地理情報には，どの都道府県に属するか曖昧な地理情報がある。例として，書籍「秋の鬼怒沼」における地名「日光」が挙げられる。この書籍における「日光」は，最も適切であると思われる「栃木県日光市日光」の他に「栃木県日光市」，「福井県福井市日光」，「愛知県津島市日光」がある。このような曖昧性のある地理情報では，誤ったジオコーディングが行われる可能性がある。

そのため本研究では，地理情報のジオコーディングに，書籍データにおける都道府県の出現頻度を用いる手法を提案する。まず，書籍データ中のすべての地理情報の中で曖昧性のない地理情報に対して，対応する“名称”があるか判定を行う。判定には，図 4 のように“名称”をスペースで分割し，“名称”の右端の文字列と地理情報が一致するか判定する。この際，“市”や“県”などの文字が省略されている場合でも，一致していると考えられる。判定により一致とみなした場合，地理情報に対応する名称であるとし，“名称”の“緯度”，“経度”を地理情報に付与する。次に，付与した“緯度”，“経度”の“名称”における都道府県の出現頻度を求める。都道府県の出現頻度の例を表 2 に示す。曖昧性のない地理情報と同様の方法で，曖昧性のある地理情報と対応する“名称”の候補を抽出する。地理情報「日光」の場合，候補は「栃木県日光市日光」，「栃木県日光市」，「福井県福井市日光」，「愛知県津島市日光」となる。次に，書籍における都道府県の出現頻度を用いて，候補から最も適切であると思われる“名称”の“緯度”，“経度”を付与する。地理情報「日光」の場合，表 2 の出現頻度が多い都道府県から順に候補の 4 件の“名称”に含まれているか判断する。それにより，候補が「栃木県日光市日光」，「栃木県日光市」の 2 件となる。本研究では，

粒度の大きい候補を優先するため，地理情報「日光」には「栃木県日光市日光」の緯度経度が付与する。このように，提案手法では，書籍における都道府県の出現頻度を用いることにより最も適切な緯度経度の付与を可能としている。

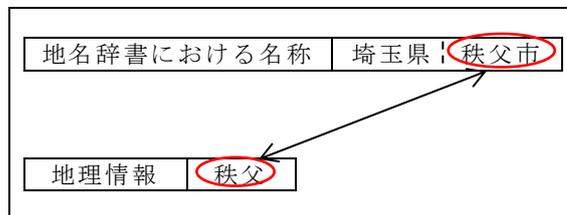


図 4.地理情報と“名称”対応付け

表 2.「秋の鬼怒沼」における都道府県の出現頻度

都道府県名	出現頻度
埼玉県	3
静岡県	1
栃木県	1
山形県	1
福島県	2
山口県	1
長野県	5
新潟県	2
秋田県	1

4.4. 書籍データのテキスト要約

本節では，書籍データのテキスト要約について述べる。書籍データにおける地理情報には，ストーリーと関連のない舞台場所が含まれている。そのため，本研究では，書籍のストーリーと関連のある舞台場所のみを提示するため，LexRank を用いたテキスト要約を行った。LexRank では，まず，文間の類似度算出する。次に，求めた類似度を基に文の重要度を算出し，文に付与する。重要度の算出には，文のグラフ表現における固有ベクトル中心性を用いて，次数の多いノードやそれに隣接するノードに重要度を付与している。本研究では，重要度の高い文に含まれる地理情報のみを用いることにより，書籍のストーリーに関連のある舞台場所のみをマッピングする。

本研究では，集約された文に含まれる粒度 3 の地理情報をストーリーに関連のある舞台場所として，地図上にマッピングしている。粒度 3 の地理情報のみを用いる理由には，粒度 3 のような具体的な場所でなければ観光場所としないと考えたからである。

5. 実験

本章では，ジオコーディングにおける提案手法の有効性とテキスト要約の有効性を確認するため行った実験と結果について述べる。5.1 節では，ジオコーディング実験，5.2 節では，テキスト要約実験について述

べる。

5.1. ジオコーディング実験

本節では、4.3 節で述べた提案手法の有効性を確認する実験について述べる。5.1.1 節では、実験方法、5.1.2 節では、実験結果、5.1.3 節では、考察について述べる。

5.1.1. 実験方法

【使用データ】

青空文庫を利用し、収集した書籍データのうち、表 3 に示す 5 件の書籍データを 4.2 節で述べたタグ付けしたデータを実験に用いる。表 3 における地理情報数とは、タグ付けにより LOCATION（地名）または ORGANIZATION（組織名）と分類された固有表現の数である。正解データには、人手で地理情報をジオコーディングしたものをを用いる。

表 3. 実験に用いる書籍データ

分野	作品名	著者名	地理情報数
紀行	秋の鬼怒沼	木暮理太郎	138
紀行	秩父の溪谷美	木暮理太郎	86
小説	駅夫日記	白柳秀湖	128
小説	或恋愛小説	芥川龍之介	15
小説	リーズ	太宰治	14

【比較手法】

- **提案手法**：曖昧性のある地理情報に対して書籍における都道府県の出現頻度を用いる手法。
- **ベースライン**：曖昧性のある地理情報に対して、候補からランダムで選択した緯度経度を付与する手法。

【評価尺度】

以下の式に示す精度を用いる。

$$\text{精度} = \frac{\text{正しくジオコーディングされた地理情報数}}{\text{ジオコーディングされた地理情報数}}$$

5.1.2. 実験結果

あ実験結果を表 4 に示す。表 4 の括弧中の数字は、左に正しくジオコーディングされた地理情報数、右にジオコーディングされた地理情報数を示している。表 4 より、書籍「秩父の溪谷美」を除いたすべてにおいて、ベースラインを上回る結果が得られた。平均をみると、0.190 向上している。

表 5 では、ジオコーディングされた地理情報の内、地名辞書に対応する“名称”が複数ある地理情報の件数と、それに対して、提案手法で正解データと一致する緯度経度を付与した件数を示している。表 4 と同様に、正解と一致した地理情報が「秩父の溪谷美」において、低い割合であることが分かる。

表 4. ジオコーディング実験結果

作品名	ベースライン	提案手法
秋の鬼怒沼	0.463(25/54)	0.574(31/54)
秩父の溪谷美	0.588(30/51)	0.569(29/51)
駅夫日記	0.397(31/78)	0.590(46/78)
或恋愛小説	0.500(3/6)	0.667(4/6)
リーズ	0.400(4/10)	0.900(9/10)
平均	0.470	0.660

表 5. 曖昧性のある地理情報における正解数

作品名	曖昧性のある地理情報数	提案手法での正解数
秋の鬼怒沼	12	6
秩父の溪谷美	22	4
駅夫日記	42	26
或恋愛小説	3	2
リーズ	7	6
平均	17.2	8.8

5.1.3. 考察

表 4 と表 5 の両方において悪い結果となった「秩父の溪谷美」では、二つの要因が考えられる。まず一つ目は、提案手法で用いる書籍における都道府県の出現頻度である。この書籍では、埼玉県秩父市の自然と他県の自然を比較する書籍であったため、都道府県の出現頻度の上位が埼玉県であった。しかし、ベースラインでは正解となった山梨県の「赤沢」に対して、提案手法では、埼玉県の「赤沢」の緯度経度を付与していた。二つ目の要因として、川や谷といった地名辞書に登録していない名称が多くあったことが挙げられる。例として抽出した地理情報に「荒川」がある。文中では、川の名前として出現した。しかし、地名辞書には川としての「荒川」は無く、地名としての「荒川」が多く登録されていた。そのため、提案手法では地名の「荒川」の緯度経度が付与されていた。このような地名辞書に登録されていないが、曖昧性のある地理情報と判定されたものが 13 件存在することが判明した。

提案手法で良い結果が得られた書籍に「駅夫日記」と「リーズ」が挙げられる。これらの書籍では、東京都が主な舞台場所であり、曖昧性のある地名である「渋谷」や「新宿」といった地理情報が多く出現した。そのため、書籍における都道府県の出現頻度の上位に東京都が登録され、多くの曖昧性のある地理情報が正解データと同じ緯度経度が付与されたと考えられる。

二つの手法において、不正解と判定したものには、CaboCha における形態素解析の誤りによるものが多く存在していた。例として「日光沢」の場合、「日」に LOCATION タグが付与されていた。そのため、「日」に「長野県東筑摩郡麻績村日」の緯度経度が付与されており、不正解と判定している。

5.2. テキスト要約実験

本節では、4.4節で述べた LexRank を用いたテキスト要約の有効性を確認する実験について述べる。5.2.1節では、実験方法、5.2.2節では、実験結果と考察について述べる。

5.2.1. 実験方法

本実験では、表3の「駅夫日記」を対象とし、ストーリーを図5に示す。「駅夫日記」に対して、4.3節の提案手法によりジオコーディングしたものを実験に用いる。そして、テキスト要約を用いない場合と用いた場合で地図上に舞台場所をマッピングする。マッピングした舞台場所が図5に示すストーリーと関連するかを人手で判定した。テキスト要約には、全693文中 LexRank により高いスコアが付与された上位207文を用いている。

目黒駅の駅夫である主人公の日常や経歴が描かれた小説。最後に転職して、九州に向かう話で終わる。

図5. 「駅夫日記」のストーリー

5.2.2. 実験結果と考察

地図上にマッピングした結果を図6、図7に示す。図6におけるピン数を表6の“要約なし”に示し、図7におけるピン数を“要約あり”に示す。実験結果の表6より、関連性のない舞台場所だけが大きく減少していることが分かる。それに比べ、関連性のある舞台場所は1件だけ減少する結果が得られた。このことより、書籍のストーリーと関連のある舞台場所のみを提示するのに、LexRank を用いたテキスト要約が有効であることが確認できる結果が得られたといえる。



図6. 要約を用いない場合のマッピング結果 (マッピング数 30件)



図7. 要約を用いたマッピング結果 (マッピング数 14件)

表6. テキスト要約実験結果

ストーリーとの関連性	要約なし	要約あり
関連性あり	8	7
関連性なし	22	7

6. おわりに

本研究では、小説や旅行記からの地理情報の自動抽出と編纂により、書籍のストーリーと関連のある舞台場所を地図上に表示するシステムの構築を行った。書籍データにおける曖昧性のない地理情報から書籍データにおける都道府県の出現頻度を用いて、曖昧性のある地理情報をジオコーディングする手法を提案した。実験結果より、提案手法は精度の平均0.660を得た。ベースラインの精度の平均0.470と比べ、精度が0.190向上させることができ、提案手法の有効性を確認できた。書籍のストーリーと関連のある舞台場所のみを提示するため、本研究ではテキスト要約を用いた。実験結果より、書籍のストーリーと関連の低い舞台場所を多く除外することができ、テキスト要約の有効性を確認できた。

参考文献

- [1] Strötgen, J. and Gertz, M., “Event-centric Search and Exploration in Document Collections”, Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, pp.223-232, 2012.
- [2] 中嶋勇人, 新妻弘崇, 太田学, “位置情報付きツイートを利用した観光ルート推薦”, 情報処理学会研究報告データベースシステム, Vol.2013-DBS-158, No.28, pp.1-6, 2013.
- [3] 河野愛樹, 中村健二, 小柳滋, “マイクロブログから抽出した地物情報と投稿間隔を考慮した位置情報推定”, 全国大会講演論文集, Vol.2011, No.1, pp.785-787, 2011.

- [4] 金木雄太, 山田剛一, 絹川博之, 中川裕志, “地名辞書を利用した地名の曖昧性解消と文書の地域分類”, 人工知能学会論文集, 2005.
- [5] 平野徹, 松尾義博, 菊井玄一郎, “地理的距離と有名度を用いた地名の曖昧性解消”, 情報処理学会全国大会講演論文集, Vol.70, No.2, pp.285-286, 2008.
- [6] 相良直樹, 砂山渡, 谷内田正彦, “サブトピックを考慮した重要文抽出による報知的要約生成”, 電子情報通信学会論文誌 D, Vol.90, No.2, pp.427-440, 2007.
- [7] Erkan, G. and Radev, D.R., “LexRank: Graph-based Lexical Centrality as Salience in Text Summarization”, Journal of Artificial Intelligence Research, Vol.22, Issue 1, pp.457-479, 2004.