

新聞記事と Web からのイベント情報の自動抽出

斉藤 隆太 石野 亜耶 難波 英嗣 竹澤 寿幸

広島市立大学大学院 情報科学研究科 〒731-3194 広島市安佐南区大塚東 3 丁目 4 番 1 号

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

あらまし 本研究では、新聞記事集合から特定のイベントについて記載された記事を自動判定し、さらにそこからイベント名、開催日時、開催地、開催施設名を抽出する手法を提案する。また、この手法を用いて新聞記事コーパスを解析し、イベントが頻繁に開催される施設名を抽出した後、その施設で開催されるイベント情報が記載された Web ページ(イベントページ)を自動判定する。提案手法の有効性を確認するため実験を行った結果、イベント記事の自動判定では精度 79.25%、再現率 67.66%が、イベント記事からの開催場所等の抽出では精度 91.51%、再現率 75.91%が、それぞれ得られた。また、イベントページの自動判定では、58.06%の精度が得られた。

キーワード 新聞記事, Web, 情報抽出, イベント情報

Automatic Extraction of Event Information from Newspaper Articles and Web Documents

Ryuta SAITO Aya ISHINO Hidetsugu NANBA and Toshiyuki TAKEZAWA

Graduate School of Information Sciences, Hiroshima City University 3-4-1 Ozuka-higashi, Asaminami-ku,
Hiroshima 731-3194 Japan

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

Abstract In this paper, we propose a method for extracting event information, such as an event name, date, a venue, and a facility name, from automatically identified newspaper articles, in which particular events are mentioned (event articles). We analyze news corpora using our method, and extract facility names from them. Then, we extract Web pages of event schedules of the facilities (event pages). To confirm the effectiveness of our method, we conducted several experiments. From the experimental results, we obtained precision of 79.25% and recall of 67.66% in automatic identification of event articles, precision of 91.51% and recall of 75.91% in automatic extraction of event information, and precision of 58.06% in automatic identification of event pages.

Keyword Newspaper, Web, Information Extraction, Event Information

1. はじめに

観光は地域における消費増加や雇用の創出など幅広い経済効果をもたらすことから、観光立国の実現を目標に掲げ、2008年10月に観光庁が設置された。その後、観光を21世紀における日本の重要な政策として位置づけた多様な取り組みが推進されている。観光を支援する媒体としては、旅行会社などが運営する観光ポータルサイトや、旅行情報雑誌「るるぶ」などの観光情報データベースなどが作成・公開されている。観光情報の中でも、美術展、コンサートなどのイベントは日々開催されており、次々に新しい情報が発信される。そのため、有益な情報提供の支援を行うためには、随

時追加されるイベント情報を新たに収集し、データベースを更新することが必要不可欠である。しかし、既存のデータベースは人手でイベント情報を抽出し、整理、保守するため、非常に時間とコストがかかる。

そこで、本研究では、低コストでのイベント情報データベースを生成し、観光客などに有益な情報提供を行うためのイベント情報検索システムの構築を目標とする。そこで我々はイベント情報の情報源として新聞記事と Web に注目し、新聞記事と Web からイベント情報の抽出を行う。まず、新聞記事について、新聞記事からイベント記事を収集し、イベント項目情報を抽出する手法を提案する。次に、Web について、Web 上

イベント検索 from Google News

① イベント名 場所 ②

③

イベント開催施設等 検索結果44件

埼玉県立近代美術館	埼玉県立博物館	埼玉工業専門学校	埼玉ベルエポック製菓専門学校	埼玉福祉専門学校	埼玉コンピュータ&医療事務専門学校
埼玉県立東松山養護学校	埼玉県立川越養護学校	埼玉県立駒西養護学校	埼玉県立越谷西養護学校	埼玉県立大宮北養護学校	埼玉県立和光養護学校
埼玉県立三郷養護学校	埼玉県立上尾養護学校	埼玉県立宮代養護学校	埼玉県立春日部養護学校	埼玉県立浦和養護学校	埼玉県立秩父養護学校
埼玉県立川口養護学校	埼玉県立越谷養護学校	埼玉県立大学	埼玉県平和資料館	埼玉県立歴史と民俗博物館	埼玉県立埼玉図書館
埼玉県立歴史資料館	埼玉県立川越図書館	埼玉県立浦和図書館	埼玉県立川の博物館	埼玉自動車学校	埼玉とだ自動車学校
サッポロビール埼玉工場	埼玉県警運転免許センター	埼玉労働局	埼玉医科大学附属総合医療センター	埼玉医大総合医療センター	埼玉医科大学総合医療センター
埼玉社会保険事務局	埼玉県消防学校	埼玉県看護協会	埼玉県宮彩の森入間公園	埼玉県宮大宮球場	埼玉県立公園
彩の国埼玉芸術劇場	埼玉芸術劇場				

④ 新聞記事掲載イベント 検索結果10件

イベント名	記事へのリンク	開催日	開催地	会場	ソース
「熊谷うちわ祭」	祭り蒐集品展:うちわ祭に合わせ開催熊谷で22日まで/埼玉-毎日新聞	22日まで	埼玉	同市筑波1のギャラリー「くまがや館」	毎日新聞 (2011年7月18日)
「復興チャリティー特別写真展」	東日本大震災:あすから越谷で「復興チャリティー特別写真展」/埼玉-毎日新聞	17、18の両日	越谷市	大袋商店街「大袋ギャラリーひろば」	毎日新聞 (2011年7月16日)
「11年度埼玉サイクリングフェスティバル」	埼玉サイクリングフェス:参加者を募集上尾などで10月16日/埼玉-毎日新聞	10月16日	上尾市など		毎日新聞 (2011年7月15日)
合同企業説明会	東日本大震災:埼玉で生徒就職を被災3県教諭 合同説明会に29人/埼玉-毎日新聞	6月20~24日			毎日新聞 (2011年7月14日)
第2回公演	被災障害者支援へ歌のリレー音楽療法士ら、埼玉で開始-朝日新聞	2月27日	川越市	高坂市民活動センター	朝日新聞 (2011年7月8日)

図1: イベント情報検索システムの例

でイベント情報の記載されているものとしては、Webニュースや「美術館」などのイベント特設ページが挙げられる。Webニュースからイベント項目情報を抽出する場合、新聞記事のモデルが利用できると考える。そこで、提案手法で作成した新聞記事のモデルを用い、Webニュースからイベント項目情報の自動抽出を行う。次に、イベント特設ページからイベント情報の抽出を行うため、本研究ではその第一段階として、イベント情報の記載されているWebページ（イベントページ）のURLを抽出する手法を提案する。本研究では、そのために過去のイベント記事から抽出した施設名とイベントページによく含まれるキーワードを用いる。今後は、抽出されたイベントページからも情報抽出を行い、イベント情報検索システムの構築を行う予定である。

本論文の構成は以下のとおりである。2節では、本研究で提案する手法の成果を利用したイベント情報検索システムを示し、その動作例を説明する。3節では関連研究、4節では提案手法について述べる。5節では提案手法の有効性を調べるために行った実験について述べ、6節で考察を行う。また、7節で結論を述べる。

2. システム動作例

本節では、イベント情報を提示するイベント情報検索システムについて、その動作例を紹介する。図1は、イベント情報検索システムの画面である。以下では、ある地域におけるイベント情報を検索する場合の一般的な操作手順について説明する。まず、画面上部の検索窓（図中①）に、検索したいイベント名や場所を入力する。（図1の場合、場所に「埼玉」という検索語が入力されている）。この状態で検索窓の横にある

「search」ボタン（図中②）をクリックすると、イベント検索結果の表（図中④）が表示される。イベント検索結果の表は、本研究の提案手法である新聞記事からのイベント情報抽出で作成したモデル（新聞記事からイベント記事を検出する際に用いるモデルとイベント記事からイベント項目情報を抽出する際に用いるモデル）を用いて、Webニュースからイベント記事を検出し、そのイベント記事から抽出したイベント情報から構成されている。また、記事へのリンクの欄にある見出しをクリックすることでイベント記事の本文の内容を閲覧することができる。しかし、新聞記事のモデルを用い、Webニュースからイベント情報を抽出した場合、イベント記事とイベント項目情報の取りこぼしが存在する。そこで、本研究では、より網羅的なイベント情報を取得するため、施設名に関するイベント情報の書かれたページのURLを抽出する手法を提案する。そして、この手法により抽出したURLは施設名一覧（図中③）にリンクされている。よって、施設名をクリックすることでイベントページを閲覧することができる。図1は、イベント名は指定せず、場所のみ「埼玉」と指定し、埼玉県の施設名とイベント情報が結果として表示されている。

3. 関連研究

本研究では新聞記事とWebからイベント情報を抽出している。本研究とはイベントの定義が異なるが、Sakakiら[1]は地震や台風をイベントと定義し、Twitterユーザの投稿からリアルタイムなつぶやきを取得する。そして、そのユーザの位置情報から地震の

発生地を検出を行っている。本研究と同様、未来のイベントについて、オンライン文書中から、将来のイベントに関する情報を推測する研究がある。金澤ら[2]はニュース記事の時系列中に現れる周期的なパターンを分析して、将来のイベントを観測する手法を提案している。本研究のシステムに用いるイベント情報の情報源である Web ニュースから情報を抽出し、未来情報年表を自動構築する研究がある。Web などの様々な媒体には未来を予測する記事が掲載されていることから、未来予測は、人々の普遍的欲求であると考えられる。そこで吉田ら[3]は、Web ニュースから未来情報の抽出し、未来情報年表を自動的に構築する手法を提案している。本研究では、ユーザ参加型でない低コストのデータベース生成を目指しているが、機械学習とユーザ知識を用い、イベント情報を構造化する研究がある。ウェブ上のドキュメントを機械可読なものとするためには、情報抽出を行い、構造化されたドキュメントに変換する必要がある。そこで森近ら[4]は、機械学習による抽出とシステム利用者の集合知を組み合わせた情報抽出手法を提案している。

ブログから情報抽出する関連研究がある。安村ら[5]は地理情報システムでイベントを扱うことを目的に、イベントの発生時間と発生場所の抽出を行っている。石野ら[6]は旅行ブログが観光情報を得るための有益な情報源であると考え、地域名と土産物の情報抽出を行っている。岡本ら[7]は、随時追加されるイベント情報を取り入れ、有益な情報提供の支援を行うため、地域イベント情報の抽出を行っている。吉田ら[8]は網羅性の高いイベント情報源の作成を目指し、ブログ記事と Web ページからイベント情報抽出法を提案している。いくつかのイベント名に対し、その前後のパターンを用いてイベント名を抽出している。本研究ではブログではなく新聞記事と Web からイベント情報抽出を行う。

4. イベント検索システムの構築

本研究では、イベント情報を収集する手法として、以下の 2 種類の手法を提案する。(1)の手法については 4.1 節、(2)の手法については 4.2 節で説明を行う。

- (1) 新聞記事からのイベント情報の抽出
- (2) イベントページの URL の自動判定

4.1. 新聞記事からのイベント情報の抽出

イベント情報を抽出する際の情報源として、イベント関連の情報を含む新聞記事であるイベント記事を使用する。4.1.1 節では、新聞記事からイベント記事を検出する手法について、4.1.2 節では、イベント記事から

イベント項目情報を抽出する手法について説明を行う。

4.1.1. イベント記事の検出

本節では、新聞記事からイベント記事を検出する手法について説明する。本研究では、イベント名、開催日時、開催地や開催施設名が含まれており、現在より未来に開催される娯楽に関する催しや行事が記載されている新聞記事を「イベント記事」とした。人手によりイベント記事と判定した新聞記事の例を図 2 に示す。なお、HEADLINE タグは見出し、TEXT タグは本文を表している。

```
<HEADLINE>東京ドーム・ふるさとフェア 9 3  
1月22日から3日間（社告）</HEADLINE>  
<TEXT> わが国最大規模の観光・物産展「東京ド  
ーム・ふるさとフェア’ 9 3」を、今年も一月二十  
二日から三日間、東京ドームで開催します。  
</TEXT>
```

図 2：イベント記事の例

イベント記事を検出するためには、イベント記事やイベント情報特有の言語表現を用いることが有用であると考えられる。そこで本研究では、図 2 の例のように、“フェア”や“物産展”、“開催します”などのイベント記事によく含まれる語を手掛かり語とする。また本研究では娯楽に関する催しが記載された新聞記事をイベント記事としているため、“会議”、“裁判”などのイベント記事に含まれにくいと考えられる語は不要語として収集する。機械学習にそれら手掛かり語と不要語の有無を素性として与えることでイベント記事の検出を行う。

機械学習には TinySVM を用い、手掛かり語と不要語から 39 の素性を使用した。使用した素性は大きく以下の 6 種類に分類される。

- イベント名に関する語(71 語)
- 日程に関する語“日”、“まで”など
- “参加費”などのイベントに関連する語(24 語)
- 政治・経済に関する語(35 語)
- 事件・裁判に関する語(48 語)
- その他“写真の有無”、“文の長さ”など

4.1.2. イベント項目情報の抽出

本節では、イベント記事からイベント項目情報を抽出する手法について説明を行う。本研究では、イベント項目情報として、イベント名、開催日時、開催地、開催施設名を抽出する。イベント項目情報を抽出するため、以下の 4 種類のタグを定義する。また、これらのタグをイベント記事に付与した例を図 3 に示す。

- ◆ EVENT : 開催イベント名
- ◆ DATE : イベントの開催日時
- ◆ ADDRESS : イベントの開催地
- ◆ LOCATION : イベントの開催施設名

```
<HEADLINE><EVENT>東京ドーム・ふるさとフェア93</EVENT> <DATE>1月22日から3日間</DATE> (社告) </HEADLINE>
<TEXT> わが国最大規模の<EVENT>観光・物産展「東京ドーム・ふるさとフェア'93」</EVENT>を、今年も<DATE>一月二十二日から三日間</DATE>、<LOCATION>東京ドーム</LOCATION>で開催します</TEXT>
```

図3: 人手によりタグを付与したイベント記事の例

イベント項目情報を抽出するためには、抽出対象の情報特有の言語表現を用いることが有用であると考えられる。そこで本研究ではイベント項目情報抽出において、それぞれのタグについての手掛かり語を収集し、それらの有無を素性とした機械学習を行う。イベント項目情報の抽出では機械学習として CRF を用いた。CRF 基本手法は与えられた文に含まれる語を分類するのに用いた。素性とタグは以下のように CRF に与える。

- (1) ターゲットとなる単語から、CRF に与える前後の単語数 k
- (2) ターゲットとなる単語の前に存在する、ターゲットからの距離が k 以内に現れる単語
- (3) ターゲットとなる単語の後に存在する、ターゲットからの距離が k 以内に現れる単語

我々は予備実験の結果から、 $k=3$ と定めた。また、機械学習には以下の 13 種類の素性を使用した。

- 単語
- 品詞
- EVENT : “展”, “ライブ”などイベントの手掛かり語(40 語)
- EVENT_2 : EVENT の前後に出現しやすい語
- DATE : “日程”, “日時”など開催日時の手掛かり語(63 語)
- DATE_2 : DATE の前後に出現しやすい語
- ADDRESS : “県”, “市”など開催地の手掛かり語(97 語)
- LOCATION : “ステージ”, “館”など開催施設名の手掛かり語(72 語)
- ADDRESS_LOCATION : ADDRESS と LOCATION の前後に出現しやすい語
- NUMBER : 全角数字, 漢数字
- KAISAI : “開催”, “行う”などの語

- KAKKO : 括弧 “ [”, “ < ”
- KAKKO : 括弧 “ () ”

4.2. イベントページの URL の自動判定

イベント情報検索システムを構築するにあたり、まず、新聞記事から作成したモデルを用い、Web ニュースからイベント項目情報の自動抽出を行う。しかし、新聞記事のモデルを用いるため、Web ニュースを情報源として用いた場合、新聞記事に出現しないイベント項目情報の抽出は困難であると考えられる。そこで、本研究では、新聞記事のモデルによる Web ニュースからイベント項目情報の自動抽出では取得できないイベント情報を補うため、Web からイベントページの検出を行う。過去のイベント記事から抽出された頻度の高い施設名にはイベント情報を含む Web ページが存在していると考えられる。そこで、本研究では、イベントページによく含まれると考えられる“イベント”, “スケジュール”, “開催”, “展”, “予定”, “大会”, “コンサート”などのキーワードと、イベント記事から抽出された施設名を用いて、イベントページの URL を取得する。イベント記事から収集した施設名のうち上位 5 件を表 1 に示す。我々は、それぞれの施設名とキーワード 1 つを組とし、順に検索を行った。検索結果の上位 1 件目に出現する頻度の高い Web ページをイベントページと仮定し、そのイベントページの URL を自動判定する。検索結果の取得には Google AJAX API¹ を用いる。

表 1: 新聞記事から取得した施設名の例

施設名	取得数 (件)
東京ドーム	393
東京都美術館	208
大阪市立美術館	150
国立科学博物館	120
国立西洋美術館	117

5. 実験と結果

本研究では、イベント情報を収集する手法として、以下の 2 種類の実験を行った。(1)の実験については 5.1 節、(2)の実験については 5.2 節で説明を行う。

(1) 新聞記事からのイベント情報の抽出

(2) イベントページの URL の自動判定

5.1. 新聞記事からのイベント情報の抽出

イベント情報を抽出する際の情報源として、イベント関連の情報を含む記事であるイベント記事を使用する。5.1.1 節では、新聞記事からイベント記事を検出す

¹ <http://code.google.com/intl/ja/apis/loader/>

る手法について、5.1.2節では、イベント記事からイベント項目情報を抽出する手法について実験を行う。

5.1.1. イベント記事の検出

イベント記事の検出には、読売新聞、朝日新聞、毎日新聞、日経新聞の4種類の新聞記事を使用した。これらの4種類の新聞記事から“開催”という単語が含まれている新聞記事2052件を抽出し、人手によりイベント記事と判定した結果を機械学習に用いる。機械学習にはTinySVMを用い、2次の多項式カーネルを使用して4分割交差検定を行った。また、精度と再現率を用いて評価を行った。実験結果を表2に示す。

表2：イベント記事の検出結果

精度 (%)	再現率 (%)
79.25	67.66

5.1.2. イベント項目情報の抽出

イベント項目情報の抽出では、人手により4種類のタグを付与したイベント記事416件を使用した。人手により付与したタグの数を表3に示す。機械学習にはCRFを用い、4分割交差検定を行った。また、精度、再現率を用いて評価を行った。イベント項目情報の抽出結果を表4に示す。

表3：人手により付与したタグの数

EVENT	693
DATE	657
ADDRESS	658
LOCATION	464

表4：イベント項目情報の抽出結果

	精度(%)	再現率(%)
EVENT	91.23	67.32
DATE	96.75	85.45
ADDRESS	90.75	81.59
LOCATION	87.32	69.26
平均	91.51	75.91

5.2. イベントページのURLの自動判定

イベントページのURLの取得にはイベント記事から収集した施設名62件を使用した。4.2節の手法によりURL62件を選出し、イベントページかどうかを手手により判定を行った。その結果、36件がイベントページであったことから、精度58.06%を算出した。

6. 考察

6.1. イベント記事の検出結果からの考察

本節では、機械が誤って検出を行った新聞記事について考察する。まず、精度低下の原因について考察を行う。誤ってイベント記事と検出された新聞記事には、イベント記事に含まれやすい手掛かり語が多く含まれていた。その例としては研究会や懇親会などの内容が書かれた新聞記事であった。

次に、再現率低下の原因について考察を行う。機会が再現できなかった新聞記事には、含まれている手掛かり語の数が少ない、イベント記事に含まれにくい手掛かり語を多く含んでいるといった特徴があった。よって、イベント記事と判定できる特徴が得られないため、機械が再現できなかったと考えられる。

6.2. イベント項目情報の抽出結果からの考察

イベント項目情報の抽出結果について、(1)機械が誤って検出した例と、(2)機械が再現できなかった例について分析を行う。

(1) 機械が誤って検出した例

人手ではタグをつけなかったが、機械が誤ってタグを付与した例について分析を行った。分析の結果、精度低下の原因は主に、(i)過去のイベントに関する情報と(ii)手掛かり語の問題の2種類である。

(i)について、誤って検出した例を表5に示す。表において、下線部分が検出できなかったタグの部分である。本研究では、新聞記事の日付などから判断し、未来に開催されるイベント情報を対象としている。表5の例では、“開かれた”という過去を表わす表現から、“シンポジウム”が過去に開催されたイベントと考えられるので誤りと判断する。本研究では、タグとその前後の手掛かり語を用いて情報を抽出している。よって、開催されるイベントと過去のイベントとの区別ができず、誤って検出してしまったと考えられる。よって、過去を表す表現を素性に追加することで改善できると考えられる。

表5：検出誤りの例

EVENT	<u>シンポジウム「食卓からアジアを見る」</u> が十二日、世田谷区で開かれた。
-------	---

(ii)について、日付や地名は記事中に頻出する語である。本研究ではDATE、ADDRESSの手掛かり語として“日”や“市”などを用いているため、イベントに関する日付や地名以外にも検出されてしまったと考える。そのため、システムを構築する際には、検出されたDATEやADDRESSの周囲にEVENTが存在しなければ

ば取り除くなどの処理を行い誤って検出されたものを取り除く必要があると考える。

(2) 機械が再現できなかった例

人手ではタグをつけたが、機械が再現できなかった例について分析を行った。分析の結果、再現率低下の原因は主に手掛かり語の不足であった。手掛かり語の不足により再現できなかった例を表 6 に示す。表において、太字がそれぞれのタグの手掛かり語、下線部分が再現できなかった部分である。

表 6: 再現できなかったタグの例

EVENT	歌人、 <u>原阿佐緒展</u> では、
DATE	「ふるさとフェア」 <u>22 日</u> 開幕
ADDRESS	二月一日まで、 <u>大丸・梅田</u> の十一階
LOCATION	開催場所は東京・世田谷区の津田 <u>ホール</u>

EVENT の例について、手掛かり語は“展”のみである。イベント名に人物名や地名などが含まれる場合に手掛かり語が不足し、特徴が得られないと考えられる。

DATE の例について、本研究では DATE の前後に存在する手掛かり語として“から”や“まで”など期間を表す語を用いている。したがって、例に示す日付のみでは開催日時と判定できる特徴が得られず、機械が再現できなかったと考えられる。改善方法として DATE の前後の手掛かり語の追加が挙げられる。

ADDRESS の例について、本研究では ADDRESS の手掛かり語として“県”、“市”などを用いている。しかし、“市”などが末尾につかない地名に関しては手掛かり語が存在しないため、再現できなかったと考える。

LOCATION の例について、末尾 1 語の“ホール”のみが手掛かり語である。したがって、開催場所と判定できるほどの特徴が得られず、システムが再現できなかったと考えられる。ADDRESS と LOCATION については地名や観光名所のリストを作成し、素性に加えることで改善できると考える。

6.3. イベントページの URL 自動判定結果からの考察

イベントページの URL の自動判定について、本研究では 62 件と少ない施設名での調査であり、施設名の内容も美術館などに偏っていた。本研究で自動判定した URL の傾向としては、施設名に関するオフィシャルページのトップページの URL が多く存在していた。5.3 節より、精度が 58.06% であることから美術館などの施設のトップページの半数以上にイベント情報が記載されていると考える。今後はさらに多くの施設名について調査し傾向を分析することにより、イベントページ

検出の精度の向上が期待できる。

7. 結論

本研究では、低コストでのイベント情報データベースを作成し、イベント情報検索システムの構築を行うため、自動的にイベント情報を収集する手法を提案した。まず、新聞記事がイベント記事を自動検出する手法を提案した。次に、イベント記事からイベント項目情報を表すためのタグ“EVENT”、“DATE”、“ADDRESS”、“LOCATION”を定義し、イベント項目情報を抽出する手法を提案した。さらに、システムを作成するにあたり、より網羅的なイベント情報収集のため、施設名とイベントページによく含まれると考えられる語を用い、イベントページの URL の自動判定を行った。提案手法には、それぞれの手法に関する手掛かり語を用いた機械学習を行った。実験の結果、イベント記事の自動検出は精度 79.25%、再現率 67.66%、イベント項目情報抽出では、精度 91.51%、再現率 75.91%という結果が得られた。イベントページの URL の自動判定については精度 58.06%という結果が得られた。

文 献

- [1] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, “Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors,” Proc. 18th International World Wide Web Conference (WWW2010), 2010.
- [2] 金澤 健介, Adam Jatowt, 小山 聡, 田中 克己, “Web からの将来情報の発見・分析にむけて,” 情報処理学会研究報告, データベース・システム研究会報告 2008(88), pp.325-330, 2008.
- [3] 吉田 光男, 乾 孝司, 山本 幹雄, “Web ニュースを用いた未来情報年表の自動構築,” 第 3 回 楽天研究開発シンポジウム, 2010.
- [4] 森近 憲行, 濱崎 雅弘, 亀田 堯宙, 大向 一輝, 武田 英明, “機械学習とユーザ知識を用いたイベント情報の構造化,” 人工知能学会全国大会(第 24 回)論文集, No.1D2-3, 長崎, 2010.
- [5] 安村 祥子, 池崎 正和, 渡邊 豊英, 牛尼 剛聡, “blog マッピングを用いたイベント情報抽出,” DEWS2007, B7-10, 2007.
- [6] 石野 亜耶, 難波 英嗣, 田熊 遥, 尾崎 貴紘, 小林 大祐, 竹澤 寿幸, “旅行ブログからの観光情報の自動抽出,” 知能と情報, Vol.22, No.6, pp.667-679, 2010.
- [7] 岡本 昌之, 菊池 匡晃, “ブログからの地域イベント情報抽出,” 情報処理, Vol.51, No.1, pp.14-17, 2010.
- [8] 吉田 将人, 福原 知宏, 増田 英考, “ブログ記事と Web ページを用いたイベント情報抽出手法の提案,” 情報処理学会研究報告, デジタルドキュメント 2009(35), pp.37-44, 2009.