4W-02

地域イベント紹介 Web 開発のための情報解析システムの提案

橋本佳奈† 清原良三‡ 寺島美昭†
創価大学 理工学部†，神奈川工科大学 情報学部‡

1 はじめに

近年情報の多様化により，人々がインターネット上から情報を意思選択，決定することが困難になっている．特に地域の情報は地域特有のものもあるため，より情報の選択決定が難しい．そこで本研究ではシステム開発者の情報解析の負担軽減，時間短縮を目的とする地域イベント紹介 Web 開発のための情報解析システムを提案する．

2 研究の背景

総務省の平成 28 年の情報通信白書によると，ウェブ化（クラウド化）などで個人が情報を発信し共有することが容易になり，個人がより多くの情報や選択肢から自分の価値観に合った意思決定ができるようになった[1]．一方では，大量の選択肢や情報をどのように選び意思決定するかが問題になっており，結果として十分な情報が得られない．地域のイベント情報はその地域のホームページなどで見られるが，地域の田舎などの情報はあまり公に公表されていないものもあり，地域の全てのイベントを把握するのは難しい[2]．地域イベント紹介 Web 開発の課題は 3 点挙げられる．課題 1「情報源となるサイトの選出」，課題 2「必要な情報の判別」，課題 3「情報の更新」である．本研究では，課題を提案する 3 つの機能で解決する．

3 情報解析システムの提案

3.1 概要

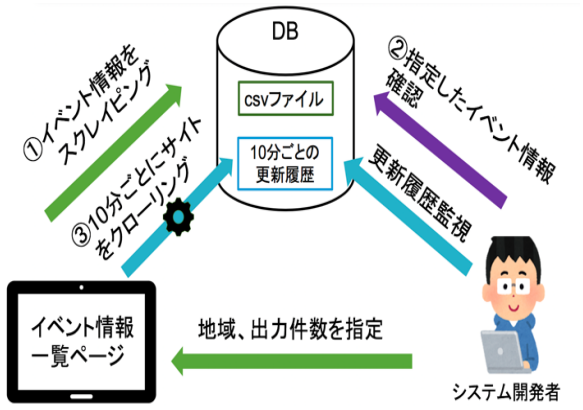

図 1 システム構成図

地域イベント紹介 Web を開発するにあたって利用するデータベースにはイベント情報，イベントの更新履歴が保存される．この開発を目的とする情報解析システムの構成を図 1 に示す．

ブラウザを自動操作できる Selenium というツールで Chrome driver を使い，①イベント情報をスクレイピングし課題 1 を解決する（サイト選出機能）．次に②指定したイベント情報を確認し課題 2 を解決する（出力内容の判別機能）．最後に③10 分ごとにサイトをクローリングし課題 3 を解決する（イベント情報のモニタリング機能）．

3.2 サイト選出機能

課題 1 に対して，イベント情報をスクレイピングすることで解決する．これは開発者が知りたい地域のイベントを Google Chrome の検索ページで検索し，スクレイピングするまでを自動化することでイベント情報サイトを探す．開発者がキーワード（地域名）を入力すると Google Chrome で指定した地域のイベント情報が検索される．検索結果（データベースに指定した地域のイベントの日付、タイトル、URL）は csv ファイルとして保存される．例として開発者が「八王子市」のイベント情報が知りたい場合を想定する．今回の研究では第一段階としてイベントの日付，イベントのタイトルをスクレイピングした．実際に Google の検索ボックスに「イベント　八王子」と入力した検索結果の一部を図 2 に示す．

図 2 「イベント＋八王子」と検索した結果の一部

図 2 で出力されたイベント情報の左上の 11 月 25 日（日）「PERSONA5～」と書かれたイベントの HTML の一部を図 3 に示す．これを見ると，日付の class は class=aXUuyd，イベント名は class=title となっている．これらは全てのイベントにおなじ class 名が付けられている．それぞれの class 名を指定することで，検索結果の全ての日付とタイトルを取得する．また，クローリングにより同時にイベントのリンクとなる URL も取得し，データベース

Copyright ©2019 Information Processing Society of Japan.
All Rights Reserved.

に保存する．

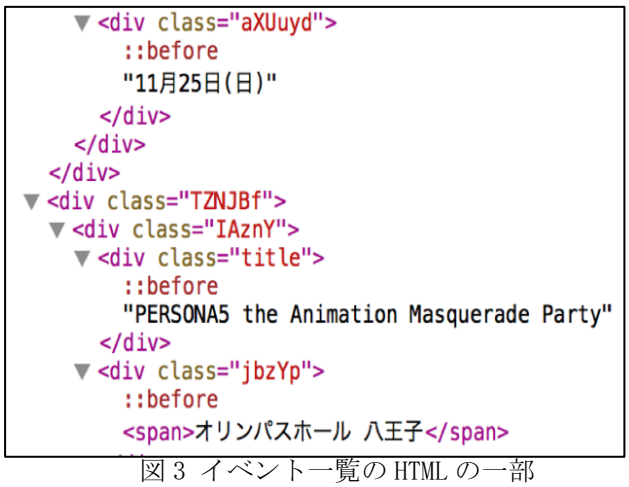

図 3 イベント一覧の HTML の一部

3.3 出力内容の判別機能

課題２に対して，開発者に適した情報を csv ファイルから出力することで課題を解決する．3.2 でデータベースに保存した csv ファイルの情報は開発者が確認することができるが，取得したデータが多いほど全てを確認するのは手間がかかる．開発者にとって必要な情報だけを出力させるために情報出力件数を指定し，開発者に表示させる．開発者が出力件数を 5 と指定した場合，図 4 の様に csv ファイルのデータから 5 件のイベント情報（イベントの日付，イベント名，イベントの URL）が出力される．出力された 5 件は日付，イベント名の五十音順，イベント名をキーワード検索，検索件数順など臨機応変にソートできる．キーワード検索については，開発者がシステムに合わせたキーワードを指定でき，データベースから検索する．

日付	イベント名	URL
11月17日(土)	町田 vs ヴェルディ	https://url/001
11月17日(土)	NBAバレエ団「くるみ割り人形」	https://url/002
11月17日(土)	MASATO HONDA in Kawasaki	https://url/003
11月18日(日)	Akishima Mayors cup in Autumn	https://url/004
11月19日(月)	Castle in the Sky	https://url/005

図 4 スクレイピング結果から 5 件出力

3.4 イベント情報のモニタリング機能

課題３に対して，イベント情報を常にモニタリングし，サイトが更新されたら情報をシステムに反映できる機能により解決する．3.2 でスクレイピングしたイベント一覧の中から開発者が気になるイベントを見つけた場合，3.3 で出力されたイベント情報の URL を指定することでサイト内の内容が変化しているかを判別する．開発者が指定した時間間隔でサイトの更新履歴をモニタリングし，更新履歴はデータベースに図 5 のように保存される．指定したサイトを 10 分おきにモニタリングすると設定した場合，更新されていなければ「更新なし」更新されていれば「更新あり」と表示される．更新されたか否かを判別する方法としては，まず指定された URL のパーマリンク（Permanent Link）をキーとする．パーマリンクとはパーマネントリンクの略で，ウェブサイトの各ページに対して個別に与えられている URL の末尾部分のことである．指定した時間間隔毎にクローラが指定されたキーに変化があるか確認し，あれば更新あり，なければ更新なしと判断する．パーマリンクがない場合は，前回クロールしたページの HTML の正規表現を比較し，変化があれば更新されたと判断し，変化がなければ更新されていないと判断する．

▼時間	▼更新履歴
10:00:00	更新なし
10:10:00	更新なし
10:20:00	更新あり
10:30:00	更新なし
10:40:00	更新なし

図 5 指定したサイトの 10 分ごとの更新履歴

4 評価と考察

サイトの選出，必要な情報の判別，情報の更新というそれぞれ 3 つの課題に対して，本研究ではサイト選出機能，出力内容の判別機能，イベント情報のモニタリング機能を用いて八王子のイベントとして地域の祭りを紹介するシステムを試作し，有効性を確認した．

5 おわりに

地域イベント紹介 Web 開発の課題を分析し，情報解析システムにより開発者の情報解析の負担軽減，時間短縮を実現した．

参考文献

[1] 総務省，”(2)情報量の増加がもたらす変化”，第 1 省，第 4 節，第 3 部，情報通信白書，平成 28 年版

[2] 佐藤他，”ソーシャルセンシングを用いた能動的な地域密着イベント情報の抽出”，一般社団法人人工知能学会，人工知能学会全国大会論文集 JSAI2017(0)，3021-3021，2017

Copyright ©2019 Information Processing Society of Japan.
All Rights Reserved.