

予感テーブル：マイクロブログのイベント情報を 利用した著名人との出会い支援システム

藤沢 和哉 元良 龍太郎 安村 通晃

{fujikzy,mtrrr,yasumura}@sfc.keio.ac.jp
慶應義塾大学 政策・メディア研究科

概要

成長のきっかけを得るために、活躍する著名人との出会いを求める人が多い。現在、トークショーや講演会の告知など著名人と出会うための情報は Web 上の各所に散らばっており、私たちはニュースサイトや口コミ、著名人本人の Web ページなどを巡回して必要な情報を探さなくてはならない。Web ページを毎日丁寧に見ることは非常に面倒で、つい見落としてしまうことも多く、問題である。本研究では、成長を望む人々が著名人と出会う機会をより多く生み出すことを目的として、Web 上に散らばる著名人の登場情報を収集・抽出し、卓上にてユーザに新着情報を通知するシステム、予感テーブルを設計・実装した。

YOKAN Table : An Intermedial System between Notable and People on Event Information of Microblog

Kazuya Fujisawa Ryutaro Matora Michiaki Yasumura
Keio University

Abstract

In order to obtain a springboard for one's growth, people often try to meet famous persons. Currently, information for meeting celebrities such as talk-show announcements are scattered in various places on internet. We have to search these information through the new sites, or by word of mouth, or even to look for the celebrities' Web sites. It is very tedious to watch these Web pages everyday carefully and sometime we may overlook the information. In this study, for assisting those people who want to grow and to meet celebrities, we designed and developed the system called YOKAN table that collects the information of celebrities's appearances, and that notifies them on top of the table as a hunch.

1 はじめに

インターネット上には著名人が出演するトークセッションや講演会のイベント情報が発信され続けている。成長のきっかけを得るため、これらの情報を元に活躍する著名人との出会いを求める人も多いが、膨大な数の Web ページの中から目的とする著名人や興味分野のイベント情報を調べるためには情報探索を効率的に行う技術が要求される。しかし、検索技術のないユーザは情報の取捨選択に時間を費やしてしまうというのが現状であり、つい見落としてしまうということも考えられる。これらの問題を解決するため、本研究では成長を望む人々に著名人と出会う機会をより多く生み出すことを目的として、多くのイベント情報が発信されるマイクロ

ブログサービスの Twitter[1] 上より著名人の出演情報を抽出・収集し、日常空間で使用されるテーブル上に新着情報を通知するシステムを設計・実装した。

一般的に検索エンジンからイベント情報を検索する場合、上位に出現する Web ページはリンク獲得数が多く得られたものやアクセス回数が多い Web ページであるため、そのイベントが既に終了していることが多い。また、トークセッションや講演会などのイベントは人数制限を設けられることが多いため、分単位で新たな情報をリアルタイムに取得することができるマイクロブログが最適であると考えた。

本研究では日付情報や URL 情報の有無に着目することで、将来開催されるイベント情報を抽出し、著名人の出演情報を閲覧できる Web サー

ビス「YOKAN」を実装した。また、将来の情報環境の中でストレスなく毎日欠かさず情報収集できるシステムを目指し、Webサービスの新着情報を通知するシステム「予感テーブル」を実装した。予感テーブルはWeb上へ誘導するための入り口となり、予感テーブル上に用意されたコースター部分にコップを置くことにより、コップの振動やコースターのカラーからイベント情報を受け取ることができるものとなっている。

2 関連研究

マイクロブログの代表として挙げられるTwitterは、2010年11月現在、日本のユーザ数が1100万人を超えるWebサービスとなっている。マイクロブログの情報を利用した研究やイベント情報に着目したサービスは数多く存在するが、その主要なものについて本章で紹介する。

藤阪らはマイクロブログから情報を発信するユーザを1つのセンサーとして捉え、マイクロブログサイトが実空間を観察するネットワークサービスであるとし、実空間で起こる様々なイベントなどをユーザからのセンシングにより理解しようと試みている[3]。社会的な分析の第一歩のため、地域イベントの影響範囲を推定することを目的とし、位置・時間・メッセージの指標を利用した分析を行っている。

岩木らはマイクロブログのユーザ同士のリンク構造に着目し、マイクロブログから有用な記事を効率的に発見するための手法を提案している[2]。ユーザと記事との近接関係が、ユーザの行動や感性の類似度あるいは単純にユーザ同士のリンク構造に表れると仮定し、特徴語の共起をもとにした感性辞書の作成や、ユーザ同士の様々なリンク構造の分析を行っている。

ことさが[5]は700以上のイベント情報掲載サイトから、独自開発のクローラで自動収集したイベント情報を閲覧できるサービスである。また、ユーザから投稿されたイベント情報についても閲覧することができる。「日時」と「場所」という要素を軸にしているため、時間に着目した検索や天気や地図などの会場周辺の関連情報も閲覧することができる。

我々の調査ではトークショーや講演会などの著名人と出会うためのイベント情報に特化したサービスは未だ存在していない。他のサービスは多くのWebサイトをクロールした後、手動でイベントの詳細情報を取得するが、本研究ではTwitterから発信される140字以内の文章からイベントの詳細情報を自動で抽出する点が異なる。

3 イベント情報の分析

我々は著名人と出会うための手段としてトークショーや講演会に参加することが手軽であり、多くの刺激が得られる場であると考えた。これらのイベント情報を通知し、著名人との出会いを支援するため、世界最大規模のマイクロブログサービスを提供するTwitterからイベント情報を収集した。

Twitterを採用した理由として、一般的な検索エンジンはリンク獲得数が多いWebページやアクセス回数が多いWebページが検索結果の上位に出現する。特定のWebページからRSSなどで定期的に情報を取得する手段がない限り、新着のWebページを発見することは難しい。しかし、Twitterではツイートと呼ばれるユーザが発言したコメントが新着の情報を持つ場合、そのコメントが他のユーザに引用され、繰り返し発言されることが多々ある。このような理由から、新着の情報をリアルタイムに収集しやすく、引用数から注目の大きさある程度把握することができるTwitterが最適であると判断し、そこから収集したイベント情報を分析した。

3.1 イベント情報の収集

2010年11月5日から2010年11月18日の2週間の間にTwitterから「トークショー」、「講演」というキーワードを検索クエリとして取得した際のコメント総数は71660件となり、平均して1日約5000件のコメントを収集できた。これらのコメント内容を確認したところ、以下の二つのタイプに大別することができた。

- 参加・特定不可能タイプのコメント
過去に行われたイベントの感想やコメント内容だけからは何についてのイベントなのかを特定できないタイプである。収集したコメントのうち、ほとんどがこれらのタイプである。
- 告知・予定など将来タイプのコメント
将来行われるイベントに対して告知や宣伝がコメントされているタイプである。参加可能なイベント情報であり、過去タイプのコメントより多くの情報が記載されているという特徴がある。

3.2 日付情報の分析

将来タイプのコメントに対してさらに分析を行ったところ、コメントに「9月23日」や「10/24」などの日付情報が含まれている場合、イ

イベントについての詳細情報が記述されていることが多いことが分かった。詳細情報とは講演会のタイトル名、ゲスト名、開催場所、公式 Web ページへのリンク URL などである。上述の、収集したコメントから日付情報を含むコメントのみを実際に抽出したところ、1 日平均約 500 件のコメントを収集できた (図 1)。

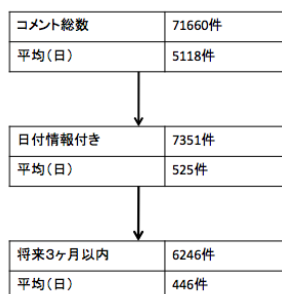


図 1: コメント分析の流れ

さらに日付情報を持つコメントのうち日付が書込み日時よりも将来3ヶ月以内であるコメント件数を調べたところ、平均して約 85% のコメントが将来3ヶ月以内の日付を持つコメントであった (図 1, 図 2)。これにより、75% から 90% の幅はあったものの、日付情報だけに着目することで将来開催されるイベントの詳細情報を取得できることが分かった。

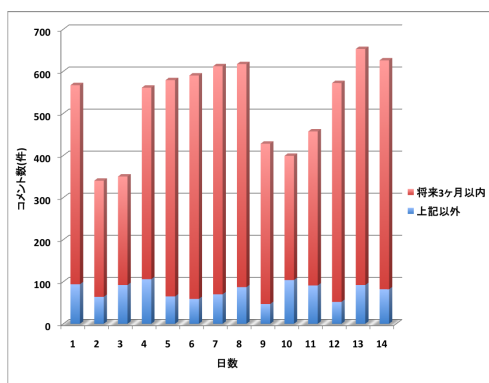


図 2: 将来情報を持つコメントの割合グラフ

以上のことから、日付情報を含むコメントを取得することで、過去に行われたイベントや特定が不可能なイベントの情報をフィルタリングして、将来開催されるイベント情報のみを取得できることが分かった。

4 設計

第 3 章で示したように、講演情報について書き込まれたコメントの中で、日付情報を含むものを収集することにより、将来開催されるイベントの詳細情報を取得できる。そこで、本研究では、それらの情報を解析した内容を一覧表示で閲覧できる Web サービス「YOKAN」を開発した。また、YOKAN の新着情報を著名人が身近にやってくることを感覚的に通知するためのシステムとして「予感テーブル」を実装した。本章では YOKAN と予感テーブルの設計について説明する。

4.1 ソフトウェア

YOKAN システムの動作手順と各処理について詳しく説明する。

4.1.1 データ収集

著名人が出演するイベント情報を検索するため、Twitter Search API を利用し、「トークショー」、「講演」というキーワードを検索クエリとしてコメントを収集した。API の取得制限回避のため、収集処理は 1 時間間隔で行った。取得したデータは以下の 6 種類である。

- 書き込まれた時間
- コメントの内容
- コメントを書き込んだユーザ名
- ユーザの URL
- ユーザのアイコン画像の URL
- コメントのパーマリンク

収集したコメントから日付情報を含むコメントのみをデータベースに格納する。このとき対象とする日付情報は「9 月 23 日」や「10/24」と書かれた月日形式のものを対象とし、月もしくは日のみの記述の場合は対象外とした。また、年情報や「明日」「三日後」「来週」などの表現で書かれたコメントについても今回の実装では対象外とした。

4.1.2 データ解析

システムの動作手順を以下に記述する。また、システム全体の流れを図3に示す。

1. 書き込み日時が最新1日分のコメントをデータベースから取得し、日付情報が現在日時よりも将来3ヶ月以内のコメントのみを抽出する。このとき、条件を3ヶ月以内とした理由として、3ヶ月目以降から4ヶ月目以内のコメント件数が数%であり、詳細情報が記述されている件数も少なかったためである。
2. Twitterでは有益な情報を発信したユーザのコメントを他のユーザ引用して発言する”リツイート”という機能が存在する。今回のようなイベント情報を対象にした場合、同じ情報が繰り返し発信されることが多々見受けられたため、第一段階として上述で抽出したコメントを日付ごとにグループ化する。
3. 次に引用されたコメントや類似する内容のコメントをクラスタリングするため、グループごとに類似度判定を行う。コメント同士の類似度判定にはベクトル空間モデルを採用した。各コメントに対して形態素解析を行い、全名詞数を特徴ベクトルの要素数、各名詞の出現回数をその要素の重みとすることで特徴ベクトルを作成する。あるコメントDの要素数をm、各要素の重みをwとしたときの特徴ベクトルは式(1)のように表される。

$$\vec{D} = (w^1, w^2, w^3, \dots, w^m) \quad (1)$$

4. ベクトル化されたコメント同士のコサイン値を類似度とし、階層的クラスタリングの最短距離法により併合を行う。ある特徴ベクトルXと特徴ベクトルYのコサイン値は式(2)により求める。類似度が閾値0.5以下になるまで併合を繰り返し、クラスタリングが終了した時点で各クラスタの一つのイベント情報として扱う。この時、各クラスタ中のコメント数を、そのイベントの注目度とする。

$$\text{sim}(X, Y) = \frac{\vec{X} \cdot \vec{Y}}{|\vec{X}| \cdot |\vec{Y}|} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \cdot \sqrt{\sum_i y_i^2}} \quad (2)$$

x_i : 特徴ベクトルXの要素 i の重み

y_i : 特徴ベクトルYの要素 i の重み

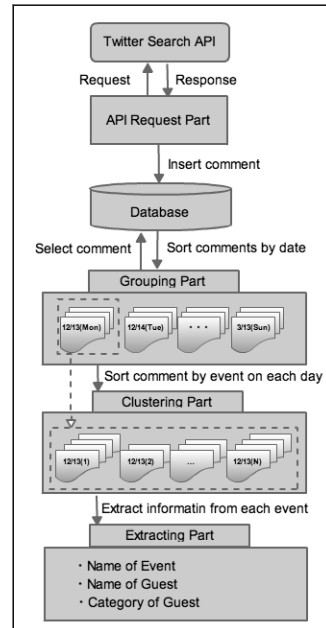


図 3: イベント情報抽出の流れ

4.1.3 イベント情報の抽出

クラスタリングにより抽出された各イベントから詳細情報を抽出する。抽出の対象となるコメントは併合されたコメント群の中で最も引用数が多かったものとし、これを基準コメントと呼ぶことにする。以下に基準コメントからのイベント情報の抽出手法について説明する。

- ゲスト名
講演を行うゲスト名を抽出するため、基準コメントを形態素解析し、品詞細分類が「人姓」と「人名」の場合のみゲストとして抽出する。また、コメント中にゲスト名が紹介される場合、「さん」「様」「氏」「先生」などの記述が多かったことから、その前の品詞が名詞の場合のみゲスト名として抽出する。ただし、後者の方法は前者の方法でゲスト名が抽出できなかった場合のみ適用する。
- イベント名
イベント名は基準コメントに含まれているURLにHTTPリクエストを行い、HTMLソース中のTitleタグの中身をイベント名として抽出する。

- カテゴリー

YOKAN では抽出したイベント情報をユーザがすべて読む負担を軽減させるため、ゲストのカテゴリー情報を表示する。カテゴリーの抽出手法として、上記のゲスト名に対して人物検索エンジン”あの人検索スパイシー”[4]より当該ゲストのタグ情報を抽出する。このサービスで利用される上位のタグ約100種類を予めシステムが決定した5つのカテゴリーに分類し、タグ情報からいずれかのカテゴリーに振分ける。

4.2 ハードウェア

予感テーブルのハードウェア構成と機構について以下で説明する。

4.2.1 ハードウェア構成

ハードウェアの構成図を図4に示す。予感テーブルのハードウェアには、市販の木製テーブル(100mm x 70mm x 70mm)を使用した。天板上に上げ底をして、内部に機構を組み込むスペースを確保した。拡張した机の内部に、DC ソレノイド (12V) を用いた振動機構および、3M 社製ポケットプロジェクタ (MPro120) を用いたカラーの投影機構を実装した。コースター部分にはスイッチを装着し、コップの重みによりスイッチのオン・オフを切り替えることができる。コップが乗っているか否かの検出および、DC ソレノイド、カラーの投影の制御には Arduino を用いた。

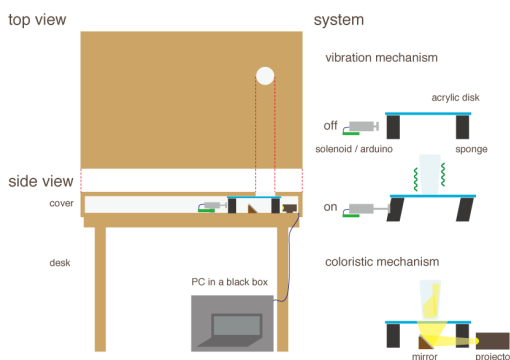


図 4: ハードウェア構成図

4.2.2 システムの連携

ソフトウェア部では抽出した講演情報から XML ファイルを作成し、この内容をパースすることで YOKAN の Web 部分に反映する仕組みとなっている。ハードウェア部は同様の XML ファイルをパースし、XML ファイルが更新された場合のみ注目度とカテゴリーの情報を一定の間隔で表示する仕組みとなっている (図 5)。

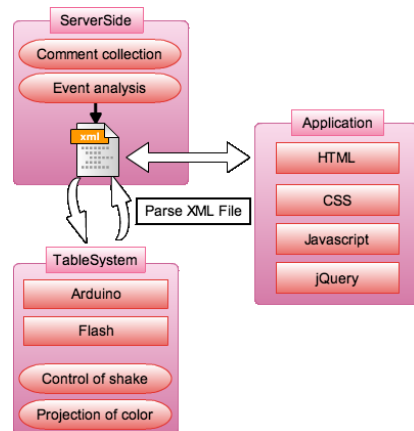


図 5: システム連携図

5 予感テーブル

予感テーブルは Web 上に散らばる著名人の登場情報を収集・抽出し、卓上にてユーザに新着情報を通知するシステムである。本システムはユーザが継続的に使用することにより効果が出やすいため、ストレスなく毎日欠かさず情報収集できるシステムであることを設計方針とした。以下、システムの Web 部分と机部分の機能について説明する。

5.1 システムの Web 部分

YOKAN は図6に示す画面構成となっている。図6の画面は2010年9月22日時点にアクセスした際の Web ページであり、9月21日に収集したコメントを解析した結果となっている。(1)ではイベントが開催される日時を表示し、日付の昇順で各イベントが一覧表示されている。(2)は講演会のゲスト名であり、抽出された人数分表示する。(3)は講演会のタイトル名であり、イベントの公式 Web ページへのリンクが貼られている。URL が存在しない場合は未表示となる。(4)はゲストのカテゴリーがカラーによって表示さ

れ、現段階で5つのカラーが存在する。現段階では大別してタレント、スポーツ、クリエイター、ビジネス、研究者の5つのカテゴリを設定している。(5)は同一のイベントに対してコメントを行ったユーザーのアイコン画像を表示している。このアイコンの数を見ることにより、どの講演会がどの程度注目されているかということを直感的に理解することができる。(6)の左のボタンをクリックすることで、(7)でイベント情報の発信元ユーザーのコメントを表示することができるため、有益な情報を発信するユーザーを発見する際の参考となる。(6)の右のボタンをクリックすることで、(7)でイベント情報を引用して発信したユーザーのコメントを表示することができるため、同じ興味・関心を持つユーザーを発見する際の参考となる。



図 6: システムの流れ

5.2 システムのテーブル部分

予感テーブルの外観を図7に示す。テーブル部分はユーザーを日常生活から Web ページへ誘導することが目的である。著名人が身近にやってくることを感覚的に通知するため、本研究では

質量の大きい物体が移動をすると、地響きで付近のものが振動するというメタファーを用いた。卓上に用意されたコースター部分にコップを置くと、コップが振動し水面が揺れ、コップの底面にカラーが表示される。新着のイベントの注目度によって揺れの大きさが、ゲストのカテゴリによってコースターのカラーが変化する仕組みとなっている。コースター部分にコップを置いておくことで Web 上の情報が順次反映されるため、ユーザーは何度もコップを操作しなくても逐次情報を受け取れるようになっている。ユーザーが朝食や夕食の際に卓上にて著名人が迫ってくる様子を感じ、通学・通勤、就業時などに細かい情報を Web ページでチェックするといった場面を想定している。



図 7: 予感テーブルの外観

6 考察

本システムでは日付情報を含むコメントが何らかのイベント情報を持つという仮定で、詳細情報の抽出を行ったが、イベントを特定できないようなコメントも存在することがあった。このような本システムではノイズとなるコメントは引用されることが少ないため、クラスタリングで抽出された各イベントが一定数以上のコメント数でなければ処理しないようにすることで対処できると考えられる。また、URLの有無や文字数の長さでフィルタリングをかけることにより、正確なイベント情報を抽出することができる。

現在、特徴ベクトルの要素数はコメント中の名詞数、各要素の重みは各名詞の出現回数としている。しかし、Twitter のコメントは 140 字以内という制限が設けられていることから、テキ

ストが短いと名詞のみの対象ではコメント同士の特徴量に大きな差がでないため、違う内容のコメントを類似するコメントだと判定してしまうことがある。そのため、名詞の中でも固有名詞だけは重みを上げるなど、品詞細分類の違いによっても重みを変化させることで、特徴量に差をだすことが必要であると考えられる。

クラスタリングを行う際、終了条件の類似度の閾値は0.5に設定している。リツイートが1つのコメントのみに対して行われている場合は閾値を0.8程度にしても問題なかった。しかし、リツイートされるコメントが2つ以上存在する場合は、各コメントのイベント情報の記載方法に違いがでるため、それを考慮した設定値となっている。この閾値が妥当かどうかは詳細な実験を行って決定する必要がある。

今後は一般的な検索エンジンを用いた場合やイベント情報ポータルサイトとの比較実験を行うと共に、多くのデータや被験者により本システムの有用性の実証していく。

7 展望

今後、予感テーブルには以下のような機能を実装していきたいと考えている。

7.1 場所情報による絞込み

開発したWebサービス「YOKAN」は日付、ゲスト名、イベント名、カテゴリー、コメント情報を表示している。それらの情報を閲覧するだけで、どのようなイベントが開催されるのかをある程度把握することができる。しかし、実際にユーザが興味を持ったイベントが存在しても、現在地より離れた地域での開催では、現地に向かうことは困難なこともある。そのため、各イベント情報について場所情報を取得し、地域条件により範囲を絞りこめるようにしたいと考えている。

7.2 パーソナライズ化

本研究で実装した「YOKAN」「予感テーブル」はプロトタイプ版のため、すべてのユーザに対して同じ情報が提供される状態となっている。本研究ではストレスなく気軽に情報を受取ることを目的としているため、ユーザが任意にカテゴリーや通知間隔を設定する機能が必要であると考えている。また、Web・テーブル共に情報を取得した後に入力できないインターフェースとなっているため、インタラクション可能な機能を追加する。

7.3 テーブル上操作

予感テーブルは注目度によりコップの振動が、イベントゲストのカテゴリーによりコップのカラーが変化する設計となっている。日常の空間の中でストレスなく感覚的に情報を受け取る手段としてこのような形となったが、予感テーブル上にどれだけの情報を提示するかは大きな課題となっている。現在の状態では気になる情報があればWebを閲覧しなければならないため、テーブルで情報を受け取った後に必ずWebで確認を行わなければならない。今後はコースター部分にイベント情報を投影する機能や、コップ操作により新着情報の切り替える機能を追加する。また、テーブル上のみで情報の受取りが完結できる仕組みも備えたいと考えている。

8 おわりに

トークイベント情報のうち、日付情報を含むコメントの約85%が将来の情報であることが分かった。本研究では、それらの情報を解析した内容を一覧表示で閲覧できるWebサービス「YOKAN」を開発した。また、著名人と出会い成長する機会を増やすことを目的として、新着情報を卓上のコップの振動にて通知するシステム「予感テーブル」を設計・実装した。これまで、著名人と出会う機会を得るためには、ニュースサイトや口コミ、著名人本人のWebページを巡回して必要な情報を探さなくてはならず面倒であった。また、毎日継続的にWebをチェックすることも面倒で、つい大切な情報を見逃してしまう。そのため、未来の情報社会では遍在した情報を目的に沿って自動で収集すること、それを日常生活内でさりげなく通知し、取捨選択の補助、見落としの防止、Webをチェックしないことへの不安感の軽減をすることが求められると考える。本研究ではそうしたコンセプトを、「著名人との出会い」を例に提案した。予感テーブルは、著名人の登場情報をWeb上から収集し一覧表示した上で、そのWebサービスへの入り口を日常生活内に設けたため、ユーザがより快適に大物との出会うきっかけを得ることができる。

参考文献

- [1] Twitter. <http://twitter.com/>
- [2] Y. Iwaki, A. JATOWT, K. Tanaka: Supporting Finding read-valuable articles in micro blogs, DEIM Forum 2009 A6-6.

- [3] T.Fujisaka, R.Lee, K.Sumiya : Estimating Influence Regions of Social Events by Geo-tagged Micro-Blogs Analysis, DEIM Forum 2010 D7-4.
- [4] あの人検索 SPYSEE. <http://spysee.jp/>
- [5] ことさが[§]. <http://cotosaga.com/>
- [6] <http://developer.yahoo.co.jp/webapi/jlp/>