

## Twitter 連携によるリアルタイム検索システムの開発

Twitter<sup>\*1</sup>社の提供するサービスは、ユーザの「今をつぶやく」新しいタイプのメディアとして世界中で注目されている。ドコモでは、モバイルでのさらなる利便性向上を目的として、ツイート検索サービスを2011年8月よりリリースした。本サービスでは、Twitterに馴染みのないユーザにも親しんでもらえるように、有名人やイベントに関する検索がされた際に一般の検索結果に混ぜてツイートを紹介する仕組みや、Twitter上でのトレンドをリアルタイムで解析する技術を開発した。本稿では、ツイート検索サービスの紹介、ならびに、サービスを支える技術について述べる。

サービス&ソリューション  
開発部

とりい だいすけ 鳥居 大祐  
よこい やすひろ 横井 靖弘

スマートコミュニケーション  
サービス部

もり ゆうじ 森 勇二

こん ゆうこ 金 祐子

### 1. まえがき

Twitterはリアルタイムにつぶやきを共有する新しいメディアとして世界中で注目を集めており、ドコモはモバイルでのさらなる利便性向上のためにTwitter社と連携し、新たなサービスの実現に取り組んでいる。

TwitterやFacebook<sup>\*2</sup>などSNSでは、ネットサービスに慣れ親しんだリテラシの高いユーザが多いとされるが、ドコモのユーザ層は幅広く、特に、普段からSNSに親しんでいない層にも広くTwitterコンテンツを親しんでもらえるサービスの提供が重要である。

そこで、有名人やイベントに関連するワードをデータベース化し、当該ワードが検索された場合に関連ツイートを通常の検索結果に混ぜて提

供する仕組みを開発した。さらに、ツイートに興味をもったユーザが一層楽しんでもらえるコンテンツを提供するために、日本ユーザ向けの話題のワード/有名人/画像、検索ワードハッシュタグ<sup>\*3</sup>をリアルタイムで解析する技術も開発した。

本稿では、今回開発したツイート検索技術やトレンド解析技術、および、2011年8月よりリリースした、iモードやスマートフォン向けに展開されている検索サービスについて解説する。

### 2. ダイレクトボックスを活用したツイート検索システム

#### 2.1 サービス概要

ダイレクトボックスは、iモードおよびスマートフォンで検索する場

合に、検索ボックス(図1(a))に入力されたキーワードに合わせ、天気や言葉の意味、映画情報など、ニーズに合わせた情報を画面に表示する部分である。

本ツイート検索システムでは、検索ボックスに入力されたキーワード(有名人やイベント)に関連するツイートを、ダイレクトボックス内にコンテンツとして表示することを実現した(図1(b))。検索キーワードに関連するTwitterアカウントを紹介することで、Twitterに馴染みのないユーザが通常の検索の中でTwitterのリアルタイムコンテンツを手軽に楽しんでもらえるよう工夫した。

さらに、ダイレクトボックス内のリンク「もっと見る」をクリックすることで、キーワードに派生するTwitterアカウント(以下、ピックア

\*1 Twitter: Twitterという名称やロゴ、Twitterバードは、アメリカ合衆国または他の国々におけるTwitter, Inc.の登録商標。

\*2 Facebook: Facebook, Inc.の商標または登録商標。

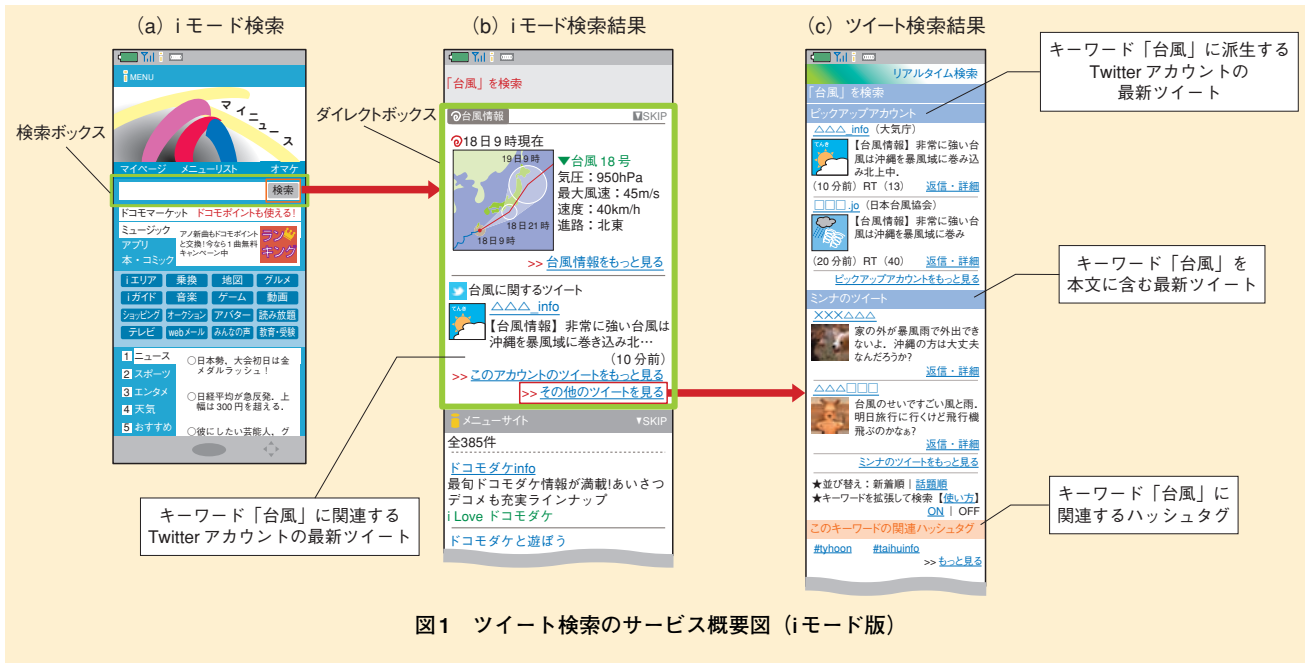


図1 ツイート検索のサービス概要図 (iモード版)

ップアカウント)のツイートや、キーワードを本文に含むツイート(以下、ミンナのツイート)、キーワードに関連するハッシュタグを閲覧することができる(図1(c))。

これらのツイートの表示においては投稿時間順だけでなく、ツイートに対する注目度を指標とした話題順での並び替えも可能である。

また、ツイートの中には、日本ではなじみのない言語や未成年などに望ましくない単語(NGワード)を含んだツイート、端末で表示できない文字コードを含んだツイートがあるため、このようなツイートは検索結果に表示されない工夫をした。

## 2.2 ツイート検索システム

図2にツイート検索システムの概要図を示す。ツイート検索は従来の

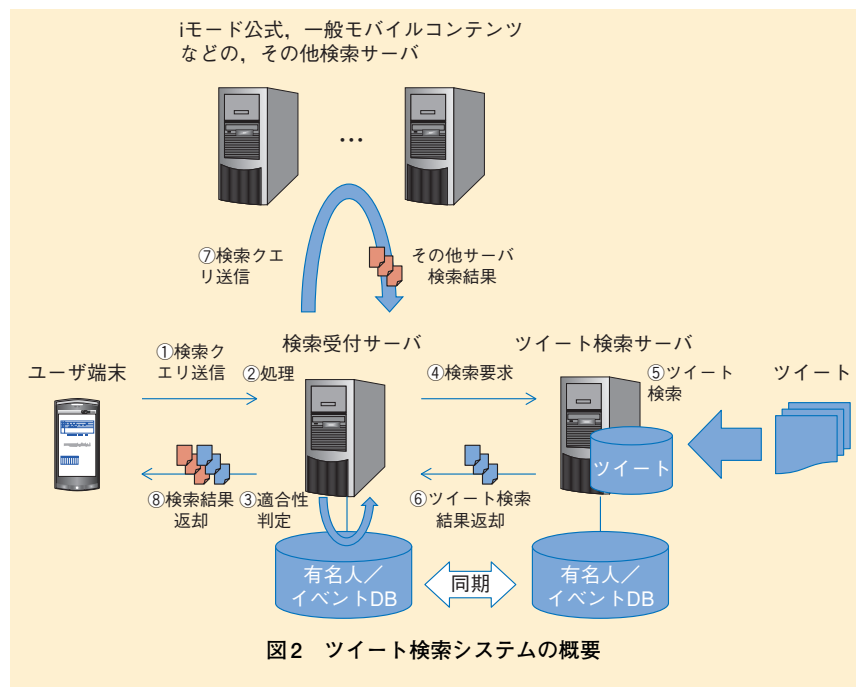


図2 ツイート検索システムの概要

検索エンジンと違い、リアルタイム性を要求されるために、Twitterからツイートを取得すると、できるだけ短い時間で検索可能となるようツイ

ート検索サーバに登録が行われる。

以下、検索が行われた際の流れについて述べる。ユーザ端末から送信された検索クエリは(図2①)、まず

\*3 ハッシュタグ：ツイートを投稿する際に「#」記号で始まる単語を付与することで、同じ話題のツイートを他のユーザが見つけやすくする機能(例) #jishin, #地震)。

検索受付サーバにて処理される (図2②). 本サーバは関連ツイートをダイレクトボックスに表示するかを判定するため, 有名人/イベントDBを用いて検索クエリとの適合性を判定する (図2③). 適合すると判断した場合, ツイート検索サーバに対してツイート検索要求を行う (図2④). ツイート検索サーバでは, 検索要求を基にツイートの検索を行う (図2⑤). 例えば, 有名人の検索要求であれば, 有名人のTwitterアカウントから最新のツイートを検索する. また, 「台風」などのイベントを表すワードであれば, 台風に関連する (複数の) アカウントから適合するツイートを検索する. 検索結果は所定のフォーマットにて検索受付サーバに返却される (図2⑥). 一方で, ユーザからの検索クエリはその他検索サーバにも送信され (図2⑦), ツイート検索の結果とともに検索結果画面を構成し (例: 図1(b)), ユーザ端末に返却される (図2⑧).

ダイレクトボックスに表示されたツイート検索結果から, 「ツイートをもっと見る」など追加情報をユーザが要求した場合や, ツイート検索結果内の検索ボックスから検索がなされた場合は, 上記で述べた検索受付サーバは経由せずに直接ツイート検索サーバに要求がされる. ツイート検索システムでは, 有名人やイベントなどのピックアップアカウントを提示するために, 検索受付サーバと同期された有名人/イベントデータベースを保有している. 検索要求

を受信した際に, 本データベースを参照してピックアップアカウントを検索結果として表示するか判定し, ヒットした場合は該当するアカウントのツイートを検索し, ピックアップアカウントとして表示する. また, ピックアップアカウント以外にも一般のツイート本文の中からキーワード検索を行い, 「ミンナのツイート」として検索結果を表示する (例: 図1(c)).

### 3. トレンド検索システム

#### 3.1 サービス概要

ツイート検索だけでなく, Twitter上での話題を提供することでさらに楽しんでもらえるようトレンド検索 (図3(b)) の提供も行っている.

トレンド検索サイトでは, Twitter上のツイート解析結果を用いることで次の4つの機能を利用できる.

##### ①話題のツイート画像

Twitter上で注目されている画像を集約して表示. さらに, 各々の画像に関連するツイートを閲覧可能 (図3(c))

##### ②話題の有名人

Twitter上で注目されている有名人をランキング表示. さらに, 各々の有名人に関連するツイートを閲覧可能 (図3(d))

##### ③話題のキーワード

Twitter上で注目されているキーワード (以下, HOTキーワード) を表示. さらに, 各々のHOTキーワードを本文を含むツイートを閲覧可能

##### ④話題のハッシュタグ

Twitter上で注目されているハッシュタグを表示. さらに, 各々のハッシュタグを含むツイートを閲覧可能

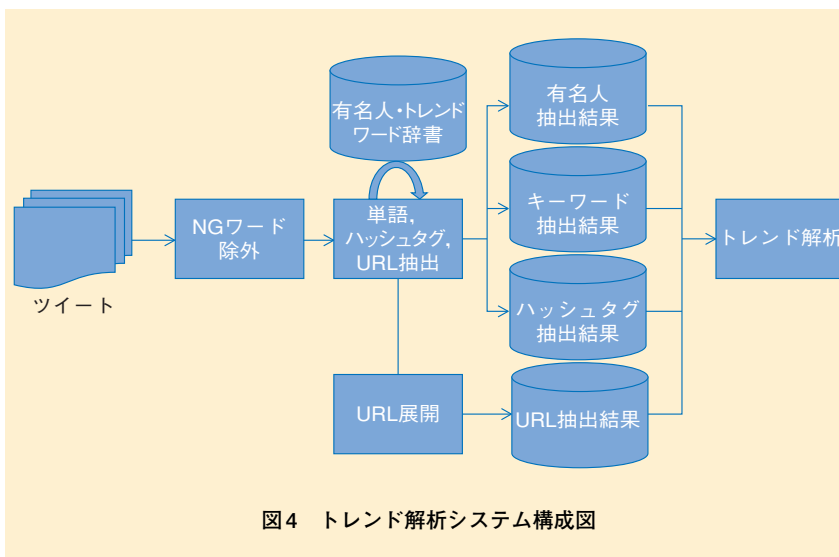
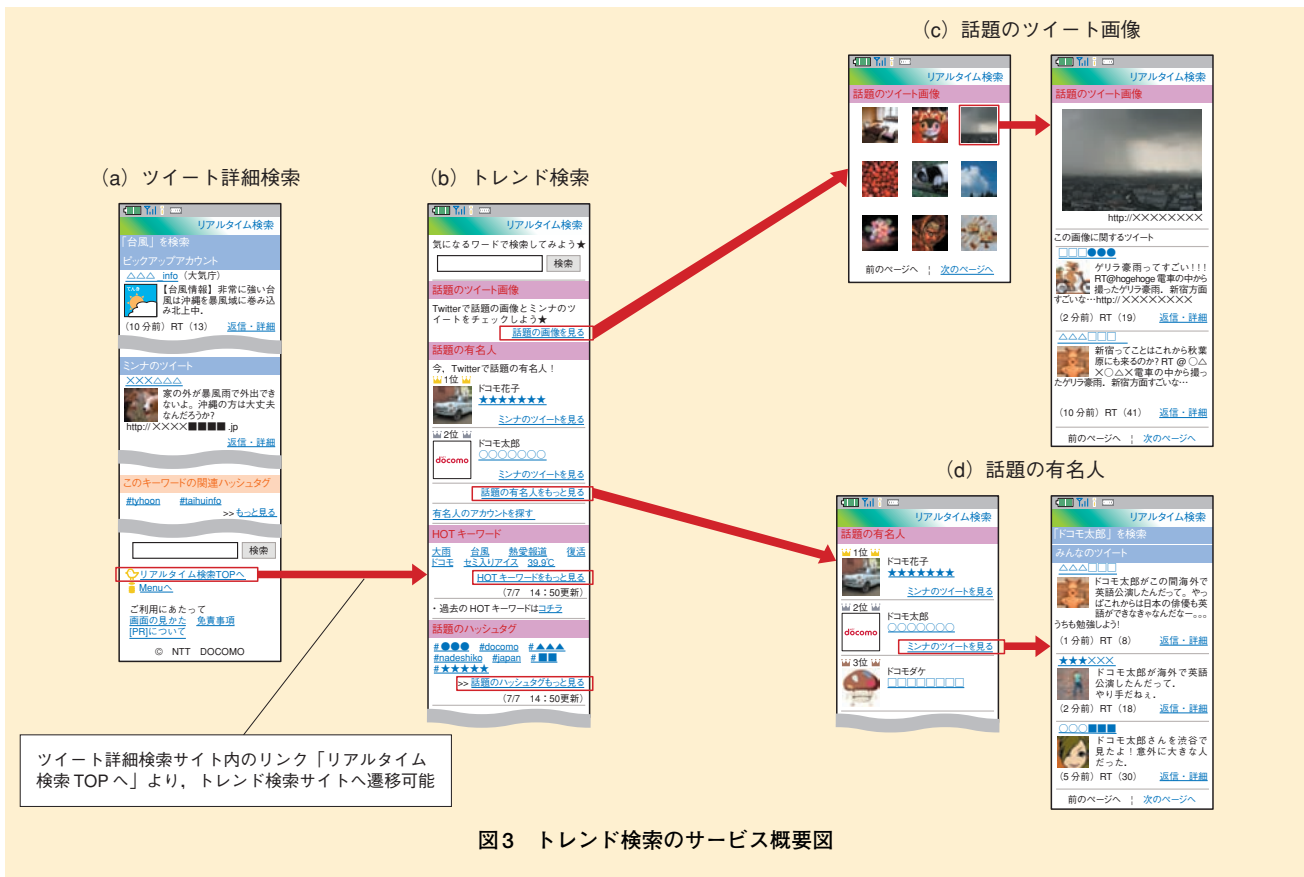
なお, HOTキーワードは前述のダイレクトボックス発動条件に適宜適用されるので, HOTキーワードが検索された場合は, 関連ツイートが検索結果として表示される. これによりユーザは, まさに世の中で起こっているリアルタイム情報をさらに知ることができる.

#### 3.2 トレンド解析システム

図4にトレンド解析システムの構成図を示す. 現在トレンドとなっている情報を提示するために, ツイート解析はリアルタイムに行われる.

以下, ツイートの解析処理の流れについて述べる. 3.1節にて紹介した4つの機能については一連の処理として説明する.

まず, 未成年などに望ましくない単語 (NGワード) を含むツイートを解析対象から除外する. 次に, 解析対象と判定されたツイートから単語, ハッシュタグ, URLを抽出する. ここで, 単語抽出にはあらかじめ用意しておいた日本ユーザ向けの有名人やトレンドワードの辞書を利用して行う. この辞書には, ツイート上でのさまざまな表現に対応するために有名人の正式名称, 呼称のほか, webをクローリングして収集した新語が含まれている. 有名人・トレンドワード辞書を用いて, ツイートを解析することにより日



行う。①話題の有名人の候補については有名人辞書内の表記揺れである正式名称、各種呼称をひとまとめにする。②ハッシュタグの抽出は#からスペースまでの一連の文字列を抽出し、辞書は用いない。③URLの抽出では、ツイート内のURLのうち、遷移先が画像共有サービスになっているものを対象とする。Twitterには投稿可能な文字列長が140文字以内という制限があることから、ツイート内のURLを短いものに置き換えるURL短縮サービスが利用されることが多い。URL短縮サービス提供者は複数存在するため、同一の画像に対して複数の短縮URLが存在し

本で話題の有名人、HOTキーワード、話題のハッシュタグの候補を抽出している。また、抽出に際して以下の処理を

得る。そこで、短縮URLについては事前に元のURLに展開しておく。

トレンドの判定は、解析の対象となる時間を決め、その時間内の出現頻度などの集計結果に基づいて行われる。前記のステップにて抽出された単語、ハッシュタグ、URLを対象に集計処理を定期的に行うことでトレンドの解析を行い、現在話題になっている画像、有名人、キーワード、ハッシュタグの提供を可能にした。

## 4. あとがき

本稿では、ドコモユーザをターゲットとして開発したリアルタイムツイート検索サービスの紹介、およびシステムの解説を行った。今回の開発では、Twitterに馴染みのないユーザにも親しんでもらえるように、有名人やイベントに関する検索がされた際に一般の検索結果に混ぜてツイートを紹介する仕組みや、Twitter上

でのトレンドをリアルタイムで解析する技術が特徴となっている。紹介したサービスは、すでにiモードおよびスマートフォンよりリリースされており、実際に利用することが可能となっている。

今後は、位置情報やその他メディアの連携などを通じて、即時に旬の情報を分かりやすく届けられるような新たなサービスおよび技術の開発を行っていく。