

実空間マイクロブログ分析による地域イベントの影響範囲推定

藤坂 達也[†] 李 龍[‡] 角谷 和俊[‡]

[†] [‡] 兵庫県立大学環境人間学部 〒670-0092 兵庫県姫路市新在家本町 1-1-12

E-mail: [†] nc06h211@stshe.u-hyogo.ac.jp, [‡] {leeryong, sumiya}@shse.u-hyogo.ac.jp

あらまし 近年、マイクロブログと呼ばれる情報メディアが普及し、短いメッセージで時空間にかかわらずユーザの状況を即時に共有することが可能となった。さらに、iPhone のようなスマートフォンから行われる場合、GPS 機能を用いた正確な位置情報をメッセージに自動で付与して発信することができるため、ジオタグ付きのメッセージが実空間に多く残されている。そこで我々は、そのようなマイクロブログを発信するユーザを 1 つセンサーと捉え、マイクロブログサイトが実空間を観察する Human Sensor Network であると定義する。これより、世界で起こる様々なイベントや出来事をマイクロブログユーザからのセンシングにより理解することが可能である。そこで、本研究ではこの考えに基づく社会的な分析の第一歩として、地域イベントの影響範囲推定を行う。我々は、移動パターンモデルを定義し、マイクロブログを位置、時間及びメッセージの指標に基づき分析し、地域イベントの影響範囲推定を行う。

キーワード マイクロブログ, Human Sensor Network, 移動パターンモデル, 影響範囲

Estimating Influence Regions of Social Events by Geo-tagged Micro-Blogs Analysis

Tatsuya FUJISAKA[†] Ryong LEE[‡] and Kazutoshi SUMIYA[‡]

[†] [‡] School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaike-honcho, Himeji, Hyogo, 670-0092, Japan

E-mail: [†] nc06h211@stshe.u-hyogo.ac.jp, [‡] { leeryong, sumiya }@shse.u-hyogo.ac.jp

Abstract Micro-blogging sites are not only a place for sharing instant personal update sharing, but also an unexplored land where we can monitor and analyze our society from a great deal of everyone's buzzes. Interestingly, recent smartphones are enabling us to easily write some micro-blogs outdoors and publish them with additional tags about automatically identified location and time. In the respect of the great number of participating users and their global distribution, we can regard such micro-blogging sites as an unprecedented human sensor network in which each person is regarded as a kind of sensor to sense the real-world events and to report their observations and opinions. In this paper, we first introduce our efforts to develop a geo-social activity monitoring system based on the micro-blogging sites by aggregating and analyzing such a novel dataset. We also present our preliminary work to find important geo-social activities with influence regions through extracting characteristic moving patterns of mobile micro-bloggers based on their publicly written spatio-temporal updates.

Keyword Micro-blog, Human Sensor Network, Movement pattern models, Influence region

1. はじめに

近年、Twitter [12]に代表されるマイクロブログは世界中の人々から注目を集めている。2009 年 6 月の Sysomos [4]の報告によると、Twitter はこの 2 年間で爆発的な成長を遂げ、ユーザ数は 1.1 億人以上となり、もっとも利用されているマイクロブログサイトとなっている。

マイクロブログサイトの大きな特徴として、ユーザは自身の最新状況を即時に短いメッセージとして発信することができる。つまり、ユーザらは周りで起こった出来事や彼ら自身の感情などを頻繁に伝える傾向にある。一般に、マイクロブログでは個人的すぎる情報が多いと考えられているが、例えばお祭りや花火大会

のような地域的なイベントが起こったとき、多くのユーザがそれに関する情報をマイクロブログサイトを通して発信する傾向にある。

また、マイクロブログはデスクトップ PC に限らず iPhone のようなスマートフォン上から専用のアプリケーション[13,15]を通して発信することができる。スマートフォンで行う場合、アウトドア環境に関する情報が多くなり、さらに GPS 機能を用いて現在の位置情報をメッセージに付与した、いわゆるジオタグ付きのマイクロブログを送ることができる。

図 1 に神戸で毎年 12 月に光の祭典として開催される「ルミナリエ」時のマイクロブログサイトから発信されたメッセージや写真を示す。



図 1 マイクロブログに反映される地域イベント

このような背景から、我々は特にジオタグされたマイクロブログに着目し、例えば何か地域的なイベントに参加しているユーザーが、ジオタグ付きのマイクロブログ（実空間マイクロブログ）をその場から発信することで、現地ではどんなことが起きているか即時に理解することができ、この考えに基づくことで世界中の様々な場所で起きている社会的な活動や自然的な現象を把握することもできると考えた。

そこで、我々はこのマイクロブログサイトを Human Sensor Network と見なしている。マイクロブログサイトにおいて、ユーザーは任意に彼らの経験、意見、感情などを書き、さらに写真やビデオも発信する。つまり、ユーザーから発信されるメッセージが地域の状況を伝えるためのセンサー的役割を担っていると思えることができる。また、人間は機械的なセンサー[3, 11]より判断する、移動する、理解するなどの優れた能力があり、それを正確に伝えることができる。

そこで、本研究ではこの考えに基づく分析として、ユーザーから発信された多くのジオタグ付きのマイクロブログを用いて社会的な活動が行われる場所を見つけ、その活動がどれくらい他の地域の人々に影響を与えるかを推定する。まず、我々はマイクロブログサイトから様々な地域に依存するマイクロブログを網羅的に取得するシステムを開発する。そして、移動パターンモデルを定義し、マイクロブログを位置、時間及びメッセージの指標に基づき分析し、地域イベントの影響範囲推定を行う。

本稿では、以下の構成で述べる。2 節で関連研究について述べる。3 節で膨大なマイクロブログの取得するための提案手法について説明する。4 節で実空間マイクロブログによる地域イベントの分析手法について述べる。5 節で実験とその結果について説明し、6 節でまとめと今後の計画について述べる。

2. 関連研究

マイクロブログはまだ発展途上のサービスであり、様々な調査が行われている。Java ら[6]や Krishnamurthy ら[7]は、Twitter の利用地域や時間におけるユーザーから発言頻度や典型的な話題について調査している。また、岩木ら[5]はマイクロブログから有用な記事の発見を行っている。これらの研究は主にメッセージの中身やリンク構造に基づく知識発見がなされているが、我々はマイクロブログでも位置や時間に着目し、特にジオタグされたマイクロブログから地域的な社会の分析を行うことで大きく異なる。

一方、守屋ら[8] はブログのような地域情報の書かれやすいテキストから地域の場所の印象や雰囲気や推定し、その結果をデジタルマップ上に表示している。こ

の研究は、社会的分析という点で類似しているが、我々はマイクロブログによる実際のユーザーらの移動パターンを分析するという点で異なる。

また、大塚ら[10]や Mohan ら[9]は実空間と Web などのネットワークの間においてどれくらい Web での行動が現実の世界に影響を与えるか発見を行う研究がなされている。我々の研究では、直接的にユーザーの行動が反映されやすいマイクロブログから地域的な側面における発見を目指している点で異なる。

3. マイクロブログ取得システム

この節では、マイクロブログサイトから多くのセンシングされたデータを取得するための手法とそのシステムの開発について述べる。図 2 にシステムの概要を示す。このシステムは、国のような大きな地域から市や町のような狭い地域まで地域範囲の粒度をオンライン地図上で操作し指定する。これにより、ユーザーの様々な要求に合わせてその地域に存在するデータを取得することが可能である。

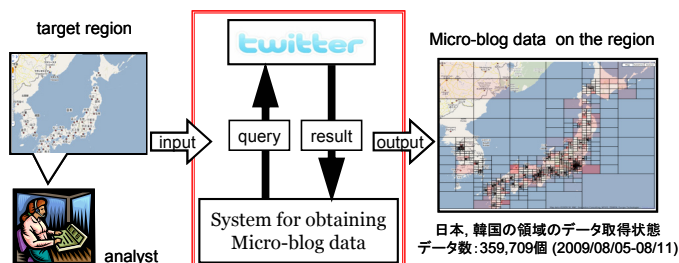


図 2 マイクロブログ取得システムの概要

具体的な取得手法について説明する。まず、Twitter API [14]を用いて、取得したい領域の中心（緯度、経度）とその半径を指定することで、その領域に属するデータを取得する。しかし、この API には制限が設けられており、1 回の問い合わせで最高 1,500 件の最新のデータしか取得することができない。つまり、指定した領域が 1,500 件以上のデータが存在する場合、それ以上のデータの取得ができない。

そこで我々は、図 3 に示した地域依存のマイクロブログを網羅的に取得する手法でデータの取得を行う。

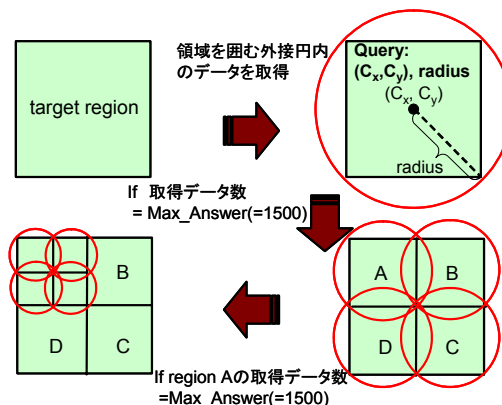


図 3 データ発生分布に適応したデータ取得手法

ここでは、まず取得範囲を指定する。そして、もし指定した領域のデータ数が 1,500 件に達した場合、システムがその領域を図 2 の A,B,C,D のように 4 分割し、

それぞれの領域で中心と半径を指定し直し、それぞれの領域に関して問い合わせを行う。さらに、もし領域Aが1,500件に達した場合は、その領域を再び4分割し問い合わせを行い、おのおの地域でデータを取得する。

これらの動作を再帰的に繰り返すことで、システムは都市部のような多くのデータが存在する地域や郊外のような少しのデータしか存在しない地域でもデータに合わせて分割するため効率よく自動でデータの取得を行い、膨大なデータを取得することを可能にした。

日本、韓国領域で行った結果の例を図2の右図に示したが、実際に日本の東京、大阪や韓国のソウル、プサンなどの都市部では小さな四角形が多く密集しており、Twitterのメッセージの量が多いことが視覚的にも理解できる。

4. 実空間マイクロブログによる地域イベントの分析手法

我々は取得した実空間マイクロブログから位置と時間によるユーザの移動傾向を分析することができると考えた。個々のユーザレベルの観点から見れば、生活場所やライフスタイルを把握することができる。また、群集レベルの観点から見れば、地域の人々の共通の移動傾向を見つけ分析することで、地域イベントや地域独特の習慣など社会的な発見ができると考えた。

そこで、本研究では個々のユーザレベルではなく群集レベルの移動に着目し、社会的な活動時に起こると考えられる基本的な2つの移動パターンモデルを提案した[1,2]。本論文では、その移動パターンモデルに基づき地域イベントを発見し、その影響範囲推定を行う。

4.1. 移動パターンモデル

我々は、社会的な活動を分析するために図4に示した2つの移動パターンモデル、集中・分散型モデルを提案した。

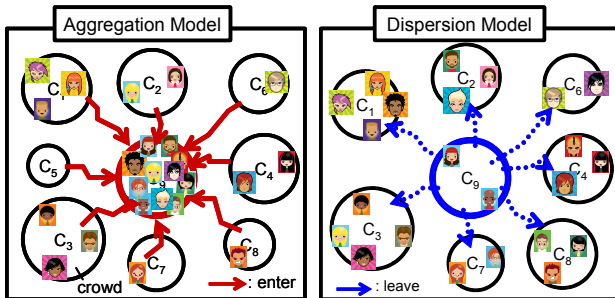


図4 移動パターンモデル

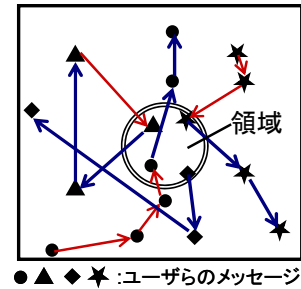
集中型モデルの概要について説明する。このモデルは、人々が移動することによりある時間に集中する場所が発生することを表すモデルである。我々は、もし人々がある時間に特定の場所に集まるとき、そこではお祭りや通勤ラッシュなどの地域的なイベントが発生している可能性があると考えた。次に、分散型モデルの概要について説明する。このモデルは、人々が移動することによりある時間に人々の分散が発生することを表すモデルである。我々は、もしある時間に普段の活動地域から分散する人が多い場合、そこでは自然災害やお盆のような帰省ラッシュなどの地域的なイベントが発生している可能性があると考えた。

我々は、この2つの移動パターンモデルに基づき地域イベントが発生している場所を発見する。

4.2. 地域イベントの発見及び影響範囲推定

地域イベントの発見とその影響範囲推定の手法について具体的に述べる。まず、地域イベント発見として、我々は人の動きが多く存在する場所ほど地域イベントが発生している可能性が高いと仮定し、人の移動の活動が激しい場所を特定することを考えた。

まず、開発したシステムより取得した実空間マイクロブログを用いて地域の分布に基づくクラスタの生成を行う。今回は、クラスタリング手法としてK-means法を用いる。次に、我々は提案した移動パターンモデルに基づく指標として図5に示した各クラスタの活動度(Activity)を定義した。



$$Activity_i = \frac{\sum_{j=1}^{\#user_i} moving\ distance(user_j)}{radius_i(km) \times \#user_i} \quad (1)$$

$radius_i$: the radius of i -th cluster
 $\#user_i$: the number of found users in i -th cluster

図5 活動度の算出

活動度では、クラスタごとに存在するメッセージから発信ユーザを特定し、ユーザらの移動履歴から移動距離をそれぞれのユーザで算出する。そして、クラスタ内のユーザらの移動距離の合計し、クラスタの大きさを考慮し上でクラスタごとの平均移動距離を算出する。これより、活動度の高い地域を見つけることで、地域的なイベントが発生している確率が高い場所を発見する。

次に、我々はメッセージ解析により、活動度の高い地域の中でも地域イベントが起こっている場所を探す。まず、下記の式(2)を用いてクラスタ内の特徴語(Characteristic Word of an Cluster)を抽出する。

$$CWC = tf \times \frac{1}{\log(df)} \quad (2)$$

ここでは、クラスタ内の出現頻度をtfとし、クラスタ間での出現頻度をdfとしtf-idfを求め名詞のみ抽出し、さらに特徴語が含まれている実際のメッセージを見ることでイベントの発見を行う。

最後に、地域イベントの影響範囲を図6のように推定する。我々はまず地域イベント発見した後で、その場所で発言されたデータから移動履歴のあるユーザ、つまり緯度・経度に違いのあるデータを持つユーザのみを分析対象とし抜粋する。次に、そのユーザがどんな地域から今の場所に到達したのか、またはその場所からどこへ離れたか移動履歴の観察を行う。次に、ユーザの移動が地域イベントに対するものかどうかを分析し、フィルタリングを行う。そして、今回はその移動の軌跡を囲む領域を影響範囲と定義し推定を行う。

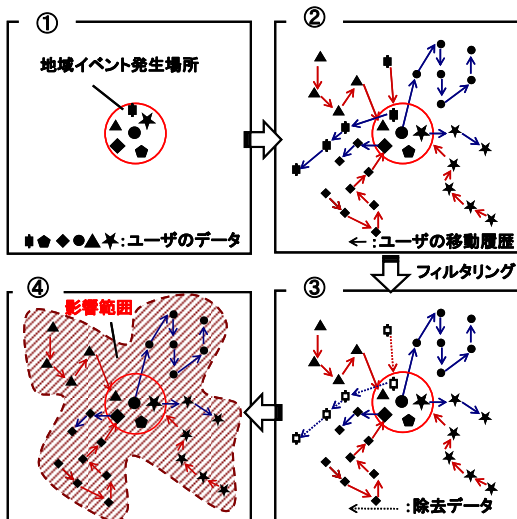


図 6 影響範囲推定手法

5. 実験と考察

我々は、実験として移動パターンモデルに基づく影響範囲推定手法の検証を行った。まず、我々が開発したマイクロブログ取得システムを用いて位置情報のある Twitter データを一週間分(2009/8/5-11)取得した。その結果、約 36 万のデータを取得できた。そして、取得したデータの中で緯度・経度の付けられたデータのみ抽出した結果、4,382 人のユーザから 128,901 個のジオタグデータを取得することができた。次に、そのデータを K-means 法を用いて緯度・経度に基づくクラスタリングした。今回は K=100 と定め行った結果、図 7 のように主要な都市部でクラスタが形成されており、特に東京や大阪周辺で多くのクラスタが存在していた。



図 7 クラスタの分布(K=100)

次に、我々はこの 100 個のクラスタすべての活動度 (Activity) を算出した。活動度の高い上位 10 件を表 1 に示す。そして、今回活動度が 9 位のお台場付近にできたクラスタのメッセージ解析を行い、降順に tf-idf 値をソートした結果、1870 個の名詞が抽出された。その上位 20 件を表 2 に示す。

この表を見てみると、灰色で塗りつぶした表のように見慣れない英語の単語が上位に多く現れた。これは、ユーザが Twitter のサイトに登録した「ユーザ名」であると考えられる。実際に、Twitter 間でユーザを指定してメッセージを書くことときに「@ユーザ名」とテキスト内を書くため、このクラスタ内のユーザが他のユーザを指定して書いたことで、ユーザ名が上位に現れたと考えられる。今回の実験では、こういったノイズを除去せずに算出したため、上位にユーザ名が現れる結果になったが、我々は正規表現を用いてこのノイズを除去することでさらに精度のよい結果が得られると考えており、今後の課題として取り組んでいく予定である。

次に、それ以外に単語を見てみると 17 位に「gundam」という特徴語が現れていた。そこで、我々はクラスタ内でガンダムに関して発信されたコメントを抽出した。その結果、図 8 に示したように写真やビデオ付きのマイクロブログも含め、このクラスタ内からそれに関する多くの発信がなされていることを発見した。

表 2 お台場付近に生成されたクラスタの tf-idf 値

rank	word	tf	idf	tf-idf
1	Tongchai	44.0	50.0	74.8
2	ryuflow	31.0	100.0	62.0
3	graphicas	23.0	100.0	46.0
4	toocheap	14.0	100.0	28.0
5	kuroishi	11.0	100.0	22.0
6	サクサク	30.0	5.3	21.6
7	Biz	21.0	7.7	18.6
8	sawa	12.0	33.3	18.3
9	arfecia	13.0	25.0	18.2
10	tyomoy	8.0	100.0	16.0
11	授乳	7.0	100.0	14.0
12	nikkatsu	7.0	100.0	14.0
13	MacBookAir	11.0	14.3	12.7
14	TechCrunch	11.0	12.5	12.1
15	kigoyama	7.0	50.0	11.9
16	ikedadaisei	7.0	50.0	11.9
17	gundam	9.0	20.0	11.7
18	roomnumber	5.0	100.0	10.0
19	jovix	5.0	100.0	10.0
20	Cyxymu	5	100	10

表 1 活動度(Activity)の高い上位 10 件のクラスタ

rank	cluster id	latitude	longitude	country	address	radius (km)	#user	#moving user	moving distance	Activity
1	18	35.64401	139.741	日本	東京都港区三田	2.6	83	19	3291.6	65.4
2	11	35.6734	139.7329	日本	東京都港区赤坂	2.1	99	23	2875.1	58.3
3	28	35.65529	139.8043	日本	東京都江東区枝川	2.7	41	13	1500.2	42.2
4	97	35.67054	139.7113	日本	東京都渋谷区神宮前	2.1	76	13	1167.9	41.9
5	77	35.69427	139.8176	日本	東京都墨田区江東橋	4.6	52	19	3593.1	41.0
6	26	35.50575	139.5703	日本	神奈川県横浜市緑区鴨居	8.3	35	12	3630.8	36.5
7	12	35.7353	139.6925	日本	東京都豊島区要町	2.7	19	4	397.6	36.4
8	96	35.66039	139.6241	日本	東京都世田谷区八幡山	3.8	45	10	1397.5	36.4
9	40	35.62043	139.7715	日本	東京都品川区東八潮	3.7	59	19	2443.9	34.8
10	81	35.6719	139.6833	日本	東京都渋谷区元代々木町	1.9	49	11	716.6	34.7

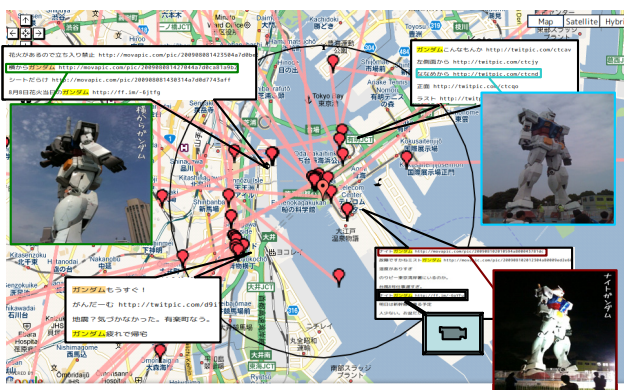


図 8 メッセージとユーザらの移動履歴の軌跡

実際に、この時期にはお台場で巨大ガンダムが展示されていた時期(2009/7/11-8/31)であり、地域イベントの発生がブログのメッセージより明らかとなった。

そこで、我々はユーザの移動履歴の軌跡を観察することで、この地域イベントの影響範囲推定を行った。実際の移動履歴を図 8 のように赤いアイコンと線で地図上に描写した結果、関東周辺の領域において移動が多かったが、九州などの遠くの地域からも移動が発生していることが発見できた。このことから、お台場の巨大ガンダムの展示はお台場周辺や関東地域に影響を与えるだけでなく、遠く離れた地域まで影響を与えている確率が高いと推定できた。

6. まとめと今後の計画

我々は、Human Sensor Network という考えた方に基づき、多くのジオタグ付のマイクロブログから地域イベントを発見し、その影響範囲推定手法を提案した。具体的には、地域依存のマイクロブログ取得システムを開発し、取得したデータから移動パターンのモデルに基づき地域イベントを発見し、推定を行った。今後の計画として、影響範囲の明確な指標を提案し、それに基づく推定を行う。

謝 辞

本研究の一部は、平成 21 年度科研費基盤研究(B)(2)「ユーザの潜在的意図を用いたレス・コンシャス情報検索基盤の構築」(課題番号: 20300039)によるものです。ここに記して謝意を表すものとします。

文 献

- [1] Tatsuya Fujisaka, Ryong Lee and Kazutoshi Sumiya, Exploring Urban Characteristics Using the Movement History of Mass Mobile Microbloggers. The Eleventh Workshop on Mobile Computing Systems and Applications (HotMobile 2010), February 2010 (to appear)
- [2] Tatsuya Fujisaka, Ryong Lee and Kazutoshi Sumiya, Discovery of User Behavior Patterns from Geo-tagged Micro-blogs 4th International Conference on Ubiquitous Information Management and Communication (ICUIMC 2010), January 2010 (to appear)
- [3] Pottie, G. J. and Kaiser, W. J. Wireless integrated network sensors. Commun. ACM, 43, 5 (2000), pp. 51-58.
- [4] "Inside Twitter: An In-depth Look Inside the Twitter World". Sysomos. 2009-06-10. Retrieved on 2009-06-23.
- [5] 岩木 祐輔, アダム ヤトフト, 田中 克己: マイクロブログにおける有用な記事の発見支援, DEIM Forum 2009 A6-6, (2009)
- [6] Akshay Java, Xiaodan Song, Tim Finin, Belle Tseng. Why we twitter: understanding micro-blogging usage and communities, Proceedings of the 9th WebKDD and 1st SNAKDD 2007 workshop on Web mining and social network analysis, San Jose, California, pp 5665, (2007).
- [7] Balachander Krishnamurthy, Phillipa Gill, Martin Arlitt. A few chirps about twitter. WOSP '08: Proceedings of the first workshop on Online social networks, pp19-24, (2008)
- [8] 守屋敬太, 佐々木史織, 清木康: 地域情報関連テキストを対象とした地域状況表示地図の動的生成方式, DEIM Forum 2009, (2009)
- [9] Prashanth Mohan, Venkata N. Padmanabhan, Ramachandran Ramjee, Nericell: rich monitoring of road and traffic conditions using mobile smartphones. Proceedings of the 6th ACM conference on Embedded network sensor systems. Raleigh, pp323-336, (2008)
- [10] 大塚 真吾, 高久 雅生, 喜連川 優, 宮崎 収兄: 女性向けフリーマガジン発行サイトにおけるユーザの行動分析, DEIM Forum 2009 A8-4, (2009)
- [11] Haenselmann, T., An FDL'ed Textbook on Sensor Networks.
http://pi4.informatik.uni-mannheim.de/~haensel/sn_book/
- [12] Twitter: <http://twitter.com/>
- [13] Twitterlator: <http://www.stone.com/Twitteltator/>
- [14] Twitter Open API: <http://apiwiki.twitter.com/>
- [15] Twiterrific:
<http://usy.jp/twitter/index.php?Twiterrific>