

センサデータとソーシャルメディアの統合による リアルタイム状況認識の可能性

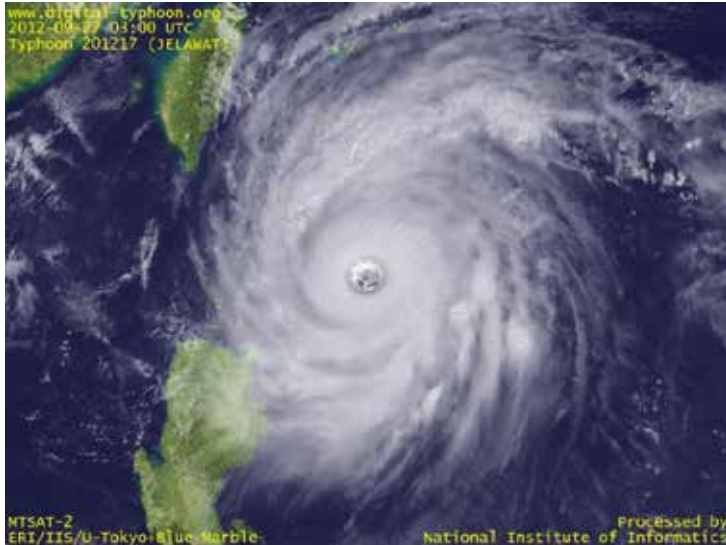
国立情報学研究所／総合研究大学院大学／

JSTさきがけ研究者

北本 朝展 (KITAMOTO Asanobu)

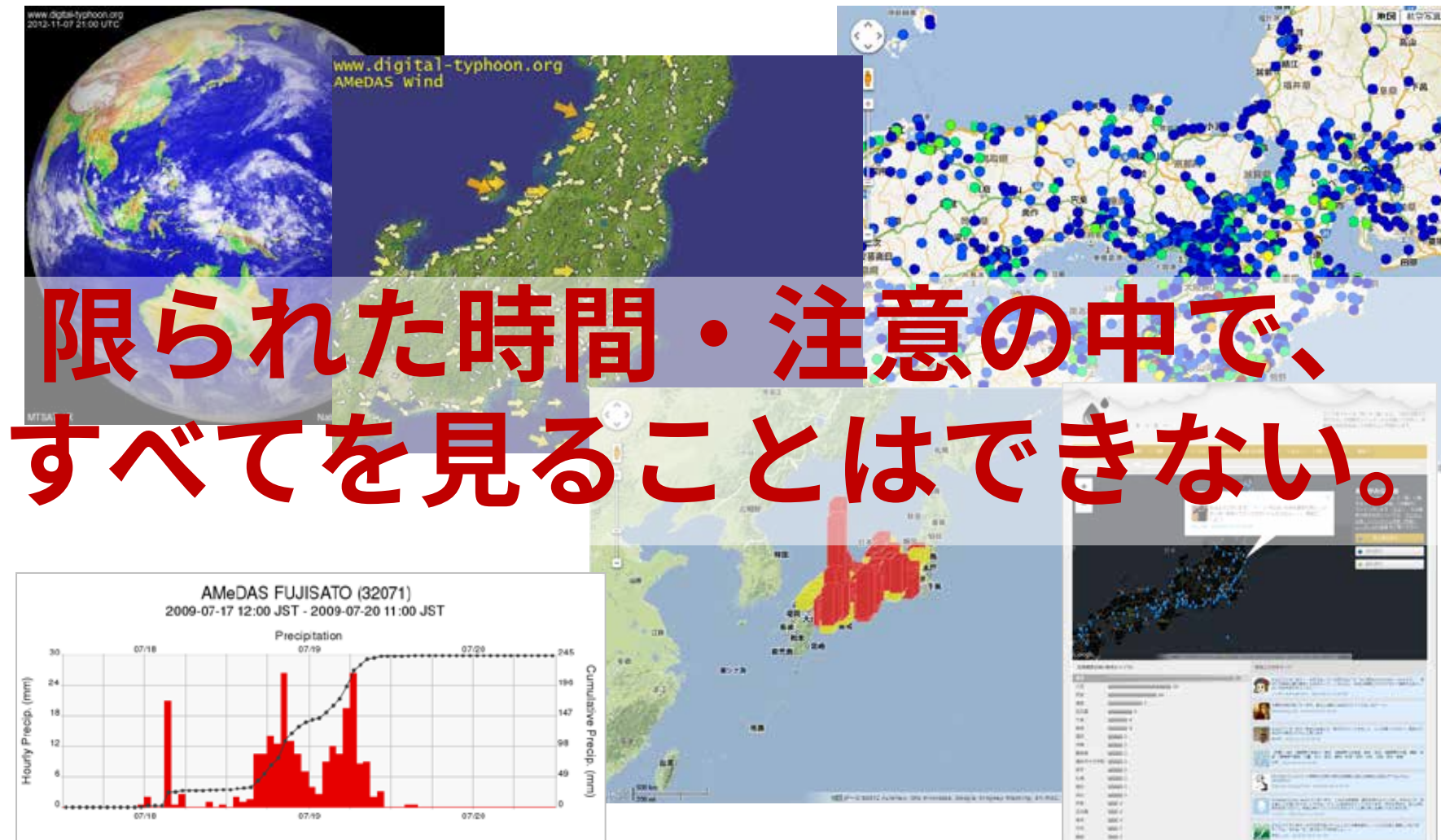
[http://agora.ex.nii.ac.jp/~ kitamoto/](http://agora.ex.nii.ac.jp/~kitamoto/)

社会の危機に対応するメディア



- 社会の危機（クライシス）には情報が爆発的に発生する。
- 必要な情報が、必要な時に、必要な人に届いているか？
- 大規模データとアルゴリズムはこの問題にどう役立つか？

クライシス時の情報爆発

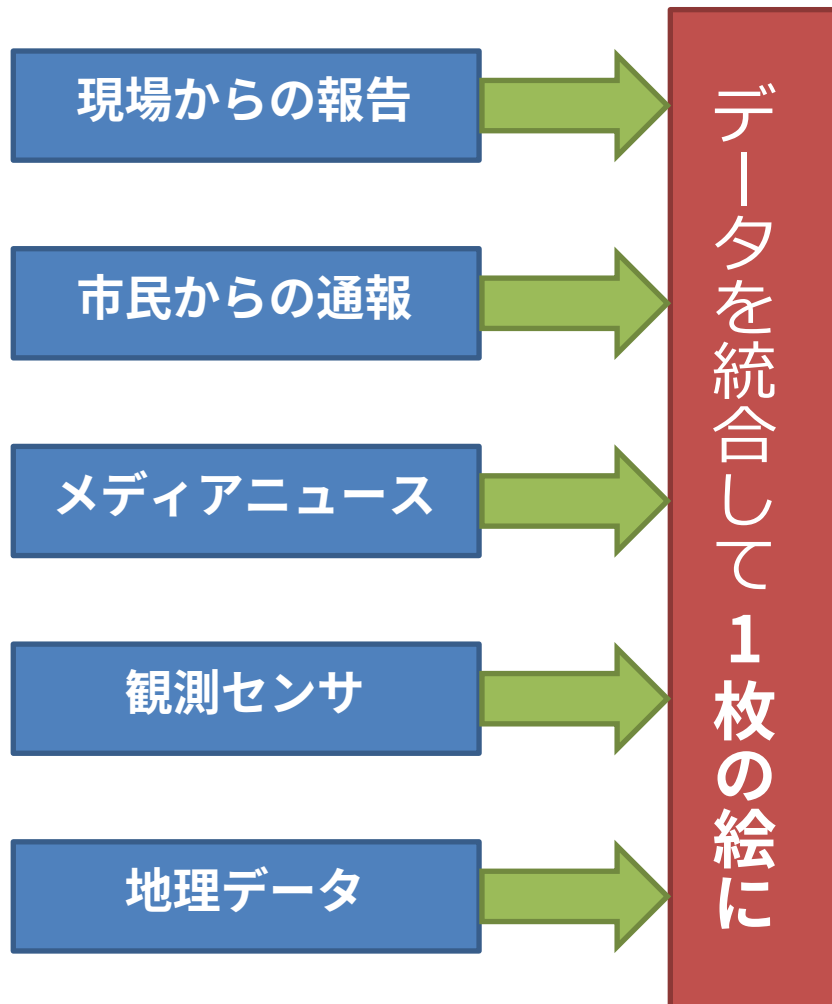


2012/11/16

ヒューマンプローブシンポジウム

3

状況認識の統一



いつ
どこで
何が
発生？

- 情報を1か所に集約し、「1枚の絵」として見せる。

降水モニタリング

手法	例
固定センサ	アメダス（気象庁）、携帯基地局センサ（NTTドコモ等）
リモートセンシング	レーダー（気象庁他）、解析雨量（気象庁・国土交通省）
専門家ネットワーク	防災関係者、協力者限定連絡網
参加型センシング	ウェザーリポート、配布センサー（ウェザーニューズ）
プローブカー	自動車ワイパー(WIDE)
ネット解析（広義のヒューマンプローブ?）	ツイッター、掲示板、SNS等のモニタリング

Human Probeの私的な分類

Human as a Vehicle (HaaV)

- 人間が搭載するセンサで観測し発信。
- 数値的・客観的なデータを取得可能。
- 位置情報はGPSで計測。
- 新たなセンサの展開に困難が伴う。

Human as a Sensor (HaaS)

- 人間の五感を記号化して入力。
- 曖昧で不安定なデータになる傾向。
- 位置情報は地名を入力。
- 既存のモバイル端末で展開できる。

竜巻目撃情報

竜巻初めて見た

2012-11-13 16:35:26

すごい!!竜巻って初めて生で見た、(^o^)/テンション上がるー♪

2012-11-13 16:36:11

松前方面で竜巻的なやつが発生しています。小さいけど…大丈夫かな…右にももう一個あったよ！

2012-11-13 16:37:31

伊予市方面で竜巻の写真が。。。こわ！

2012-11-13 16:37:52

気象庁の見解=16:29ごろ発生



https://twitter.com/pachi_mocha/status/268256260892266496/photo/1

ヒューマンプロローブの課題

1. アクセス制限等の制約のもとで、**効率的に情報を収集（クロール）**する。
2. 自然言語処理、画像処理等を用いて、**曖昧なデータから有用な情報**を得る。
3. **重要な（現地の）情報を発信するセンサ**を素早く発見し、分析する。
4. **多くの人を巻き込むインセンティブ**を設計し、**持続性の高いサービス**とする。

デジタル台風

<http://agora.ex.nii.ac.jp/digital-typhoon/>

The screenshot shows the Digital Typhoon website. The main header is blue with the title 'デジタル台風：台風画像と台風情報'. Below it, there's a navigation bar with 'ホーム', 'Earth', and 'デジタル台風'. A message in red text says 'この時間に台風情報は入っておりません。' (No typhoon information is available at this time). The right sidebar contains '台風発生数' (Typhoon occurrence count) and '最新台風情報' (Latest typhoon information). The bottom left has a '台風データベース' (Typhoon database) with search options. The bottom right shows a satellite image of the Earth with the date '2012年1月27日10時 (JST)'.

- 台風に関する「あらゆる情報」を集約するサイト。
- 最新状況の把握と過去アーカイブの検索がシームレス接続。
- 年間3000万ページビュー程度の利用。

気象現象とソーシャルメディア

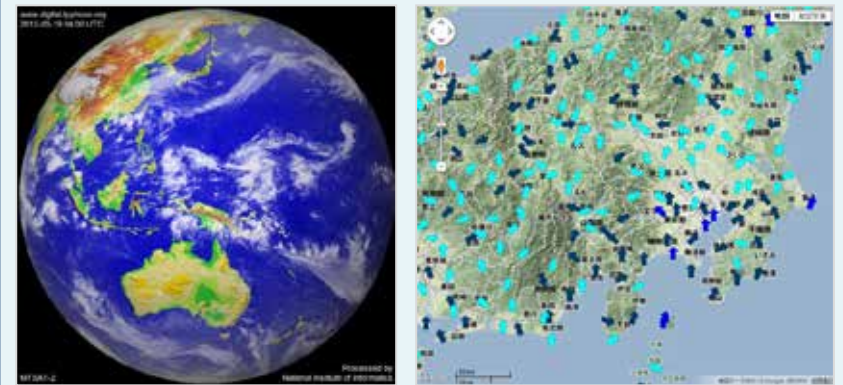
気象情報の発信



2008年4月～

速報性・共有性に優れたメディア

気象現象の解析



衛星センサ

気象センサ



人間の五感を
記号化する、
特性が未知の
センサ

人間が主体のソーシャルセンサ

ソーシャルセンサの取材方式

通報方式	対話方式	巡回方式
何か重要なことがあれば、連絡先に通報する。	問い合わせに対して、対話や情報提供を行う。	何か重要なことはないか、定期的に見て回る。
110番・119番	相談窓口	パトロール
ノイズが少なく効率的だが、参加の敷居が高い。	価値ある情報のフィードバックで動機づける。	ノイズは多いが、参加者の動機に依存しない。
ツイフーン	台風なう！	ふってきったー
2009年10月	2010年3月	2012年1月

通報方式

<http://twiphoon.eye.tc/>



- **アイフーン・ツイフーン**：現地から発信する「わたし」の台風情報。
- 利用頻度はトラックバック△、ツイッター×。
- わざわざ通報する動機が不明確？

対話方式

<http://typhoonnow.eye.tc/>

The screenshot shows the TyphoonNow website interface. At the top, there's a header with the site name in Japanese and English, and a navigation bar. Below the header, there's a section titled '台風問合せと返信 (台風201115号)' (Typhoon Inquiry and Reply (Typhoon 201115)). This section contains a list of tweets from users and the TyphoonNow account, providing real-time updates on Typhoon 201115. The tweets include location information, time, and the status of the typhoon. The interface is bilingual, with Japanese and English text visible.

- **台風なう！**：現地から聞く「あなた」の台風情報。
- ツイッターの返信機能を利用。
- ユーザの状況に応じた情報をフィードバックできれば、発展の余地あり。

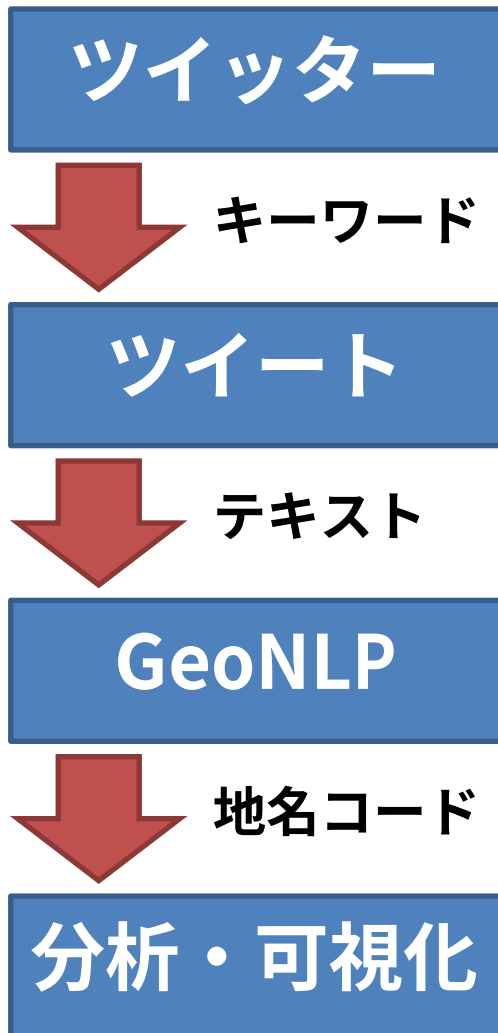
巡回方式

<http://agora.ex.nii.ac.jp/futtekitter/>



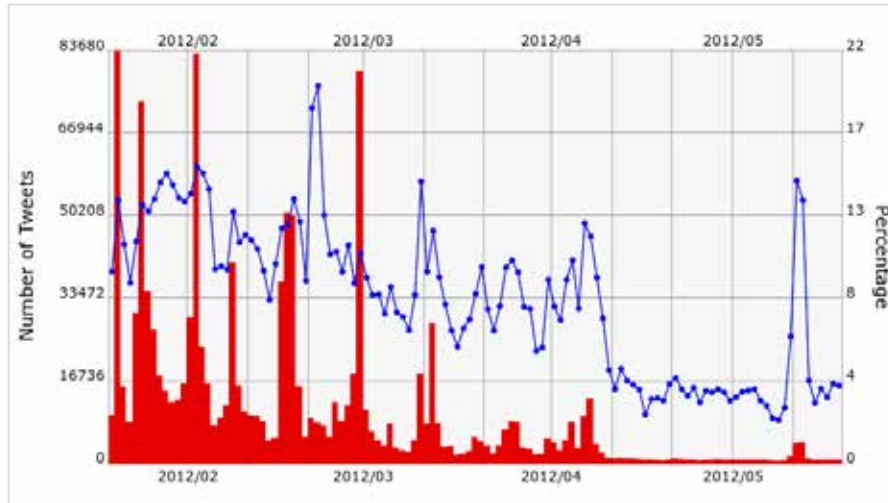
- **ふってきッター**：
雨や雪のツイートを、キーワード設定で自動的に収集。
- ツイッターのAPI
を利用。
- 大量の情報を収集
できるが、メタ
データに乏しい。

ソーシャルメディアの観測

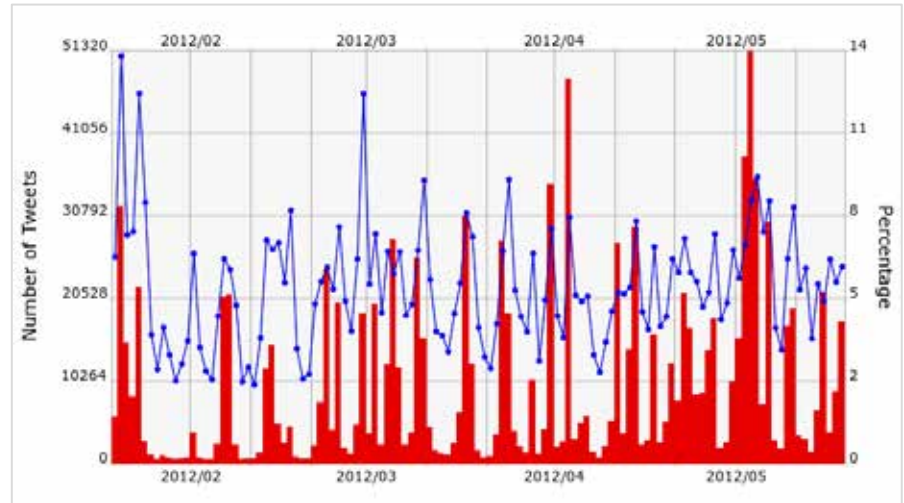


- 「雨」や「雪」等の単語を含むツイートを収集。
- **テキストに出現する地名を抽出**（ツイートのGPS情報は不使用）。
- **ふってきったー**：気象現象へのソーシャルな反応を、地図上にリアルタイム可視化（10分ごと）。

「雪」「雨」 ツイート数の変化



「雪」 ツイート数（地名含むもの）

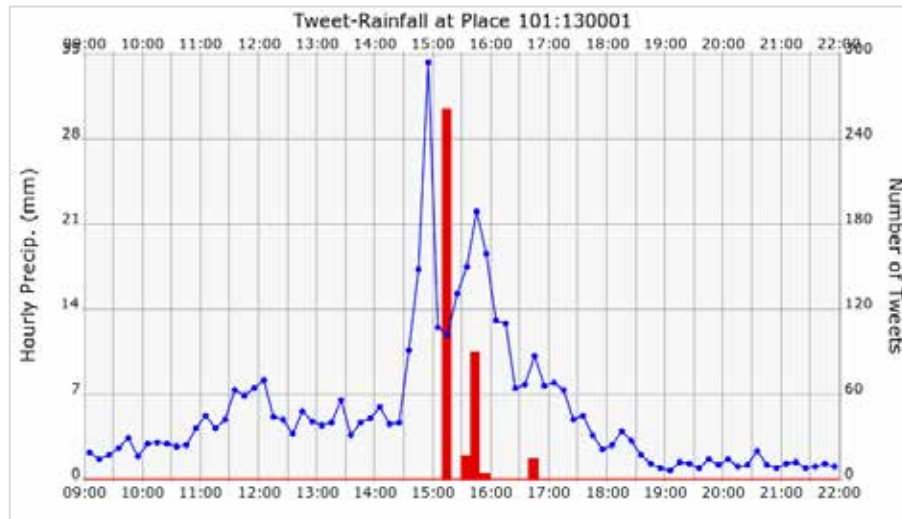


「雨」 ツイート数（地名含むもの）

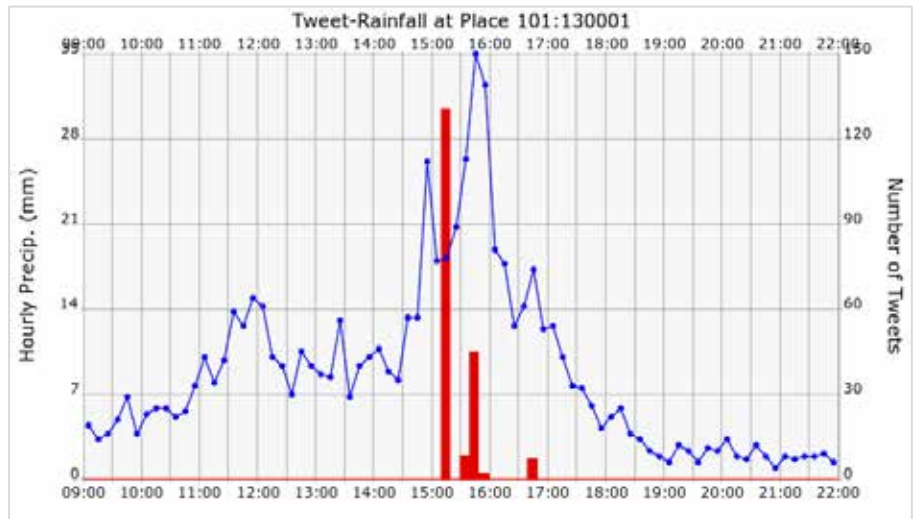
- ・ 赤は地名を含むツイート数（リツイート除く）
- ・ 青は地名を含むツイートの割合（パーセント）

- 雨や雪などの降水イベントでツイート増加。
- 地名を含むツイートは10%以下にとどまる。

降水とツイートの関係



レーダー雨量+ツイート



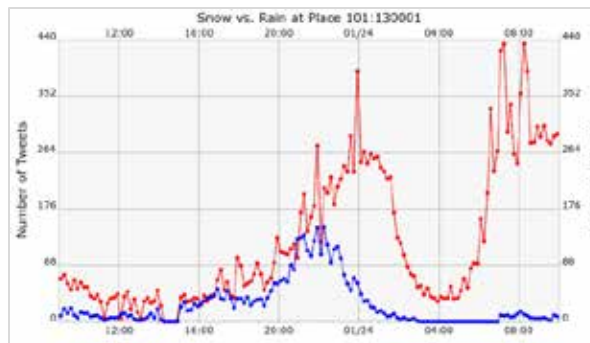
レーダー雨量+ツイート
(リツイート除く)

- 降水直前のツイート数増加は、天気予報や気象情報のリツイートが原因？

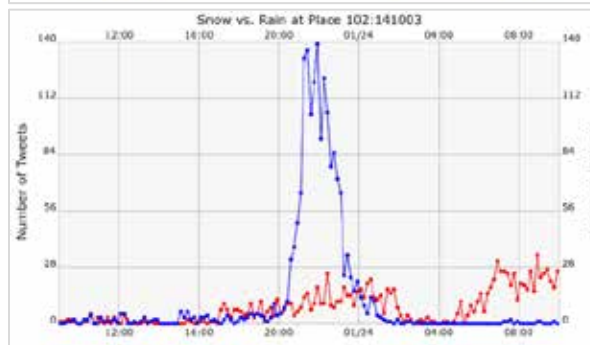
2012年1月23日の降雪



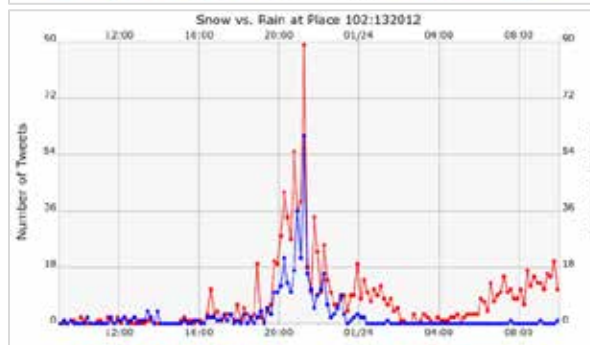
赤：雪ツイート、青：雨ツイート
(リツイート除く)



東京



横浜



八王子

ソーシャルメディアの偏り

雨ツイート数ランキング

都道府県	ツイート数
東京	359,315
大阪	119,777
神奈川	110,147
愛知	62,819
千葉	51,408
島根	1,874

雪ツイート数ランキング

都道府県	ツイート数
東京	278,476
北海道	155,986
新潟	91,220
大阪	72,923
愛知	61,215
徳島	3,012

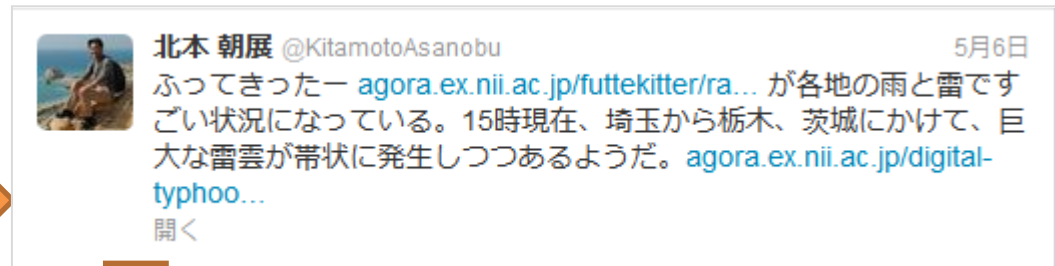
調査対象：2012年1月19日から2012年5月19日

ソーシャルメディアの検証



キーワー
ド検索

ソーシャルストリーム



GeoNLPによる地名ベ
ースのジオタギング

(time, latitude, longitude, situation)

降水状況に関するソーシャル・
科学ストリームの比較。



科学ストリーム



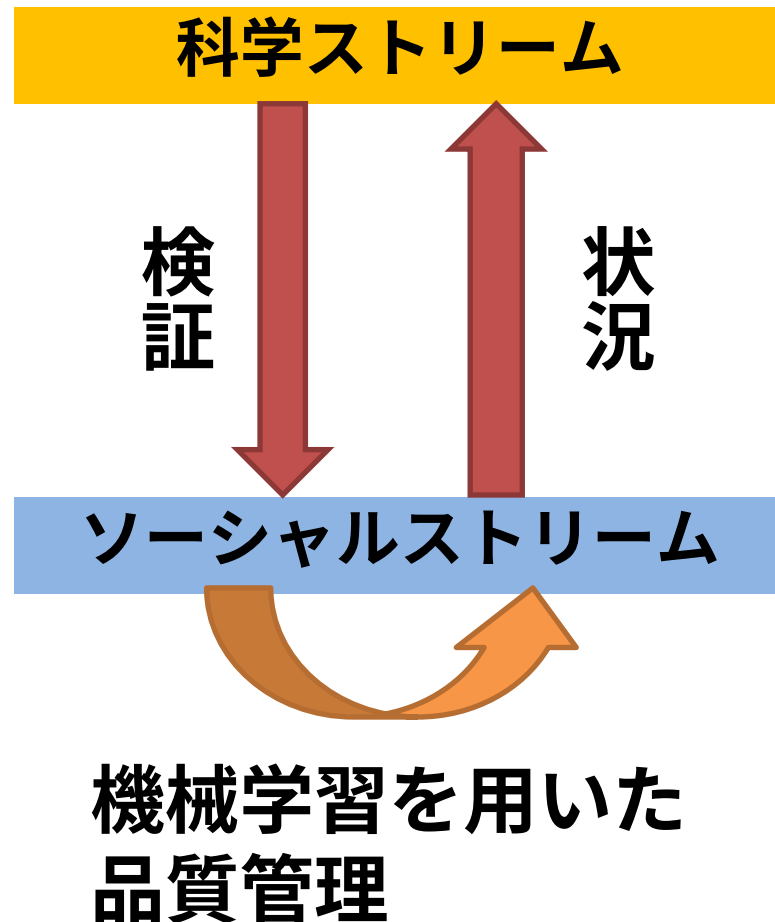
- 天気一般に比べて降水は実際に目で観察しやすいという利点がある。
- レーダーを使えば実際に降水があったかどうかを確認しやすいという利点もある。

ソーシャル／科学ストリーム



ツイート数（ソーシャルストリーム）を気象レーダー（科学ストリーム）に重ねたもの。左：リツイート除く、右：リツイート含む。

異種ストリームの相互検証



- 科学データを正解データとすれば、ツイートの分類に機械学習が活用できる？
- ソーシャルメディアデータに対する品質管理プロセスを構築したい。

GPSと比べた地名の優位性

種類	ツイート数	地名数	地名数／ツイート数
雨	3654564	4596602	1.26
雪	1738441	2207735	1.27
雷	478471	619218	1.29
台風	855120	1114163	1.30
ひょう	62369	70901	1.14



GPS情報が付与されたツイートは**全体の0.15%**しかない。



地名を一つ以上含むツイートは**全体の10%程度**はある。

- 地名を含む100万以上のツイートが1年以内に収集できた。
- キーワード検索のため、気象現象文字列を含む、気象とは無関係のツイートが混入する。
- 20-30%のツイートは複数地名を含む。

twitter 横浜町

rusty_benedict 怪しい顔月 ペネライトさん
#テレビ朝日の富川くん、昨日は新潟にいたと思ったが今日は青森・横浜町に取材かあ #報道ステーション
2月2日

LoveGreeeeeN たあ坊
横浜町があるのか!?
2月2日

shoko_n18 しょうこ
青森に横浜町ってあるんだね
2月2日

sssskkaattss 豆腐
横浜町すげえなまじか
2月2日

famous_novelist 志賀さん
青森に横浜町なんてあるのか!
2月2日

ho_ro_now ごろう
青森に横浜町ってあるんだ!
2月2日

tanoshiiko しいこ☆
青森に横浜町って町があるらしい!! おもしろー!! 青森出身のくせに知らなかった〜(๑_ω_๑)
2月2日

kurnn くりっち
青森県横浜町
2月2日

nao_0414 ナオ@旅行好き
うわー。横浜町(T_T)
2月2日

hitotsch hitopan
青森県横浜町の人とは横浜出身と言うのかしら?
2月2日

junmsize 純 (M)
青森県横浜町って「横浜出身なんだー」「へー都会だね!」詐欺出来そう
2月2日

朝日新聞 THE ASAHI SHIMBUN DIGITAL

24時刊 朝刊 You刊

トップ ニュース スポーツ エンタメ ライフ ショッピング Astand トピックス

社会 ビジネス 政治 国際 文化 サイエンス 社説 コラム 天気 交通 動画

現在位置: 朝日新聞デジタル > ニュース > 社会 > 時事通信ニュース > 記事

2012年2月2日 18時6分

雪道で数百台動けず＝青森、40キロ通行止め＝立ち往生、山形でも

[PR]

mouse computer MousePro

Windows® 7 Professional 搭載
[キャンペーン特別モデル]
39,900円～
法人PC販売開始!
■ Windows® 7 Professional 64
ビット 正規版
MousePro-152125-0116
開庫期間:
1月16日～3月20日受注分まで対応
詳細はこちら

1日午後5時ごろ、下北半島にある青森県横浜町の国道279号で大型自動車スリップし、道をふさぐ事故があった。同町によると、この事故や、大雪で道幅が狭くなるなどの影響で渋滞が発生し、車数百台が立ち往生している。県警によると、けが人の情報は入っていない。

県は午後5時、同県野辺地町有戸からむつ市大曲の約40キロの区間を通行止めとし、横浜町と協力して除雪車を現地に派遣、流氷の解氷を急いでいる。三村申吾知事は2日午前0時半、海上自衛隊大湊地方総監に対し、災害派遣を要請した。

新刊雑誌のご案内 朝日新聞 事業・サービス紹介

言葉の壁より、高い壁?

グローバル化を目指す企業の真の課題とは何だろう。
詳しくはこちら

[PR]注目情報
ビジネスの諸問題を解決!
マイクロソフトのOffice 365
を、使いこなすセミナー開催

[PR]注目情報
BMWで4つの旅をプレゼント
xDriveで北海道から沖縄まで、
BMWの最新車種を1泊2泊で体験

- 同綴異義語：一つの綴りが複数の場所に対応。
- 一意な地名解決に必要な手がかりが不足。
- 自然言語文ジョタギングシステムGeoNLPを開発。

地名情報処理の困難性

- **同綴異義語**

- 青森県の横浜と神奈川県横浜

- **品詞の曖昧性**

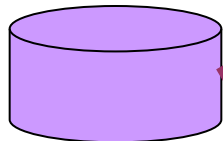
- 津波（つは）に行く → かなり難しい
- 津波を見に行く → 非常に難しい
- 風呂に行く → テキストだけでは困難？

- **表記揺れや異表記、境界の曖昧性**

- 「ケ」と「ヶ」、「ノ」と「の」
- 湯布院と由布院、福島第一原発とフクイチ

ジオタギング by GeoNLP

地名に関する
公開情報源



<http://agora.ex.nii.ac.jp/GeoNLP/>

テキスト / HTMLを
JSON-RPC APIに入力

自然言語テキスト

2012年度中にソフトウェア
と地名辞書をオープンソース
化する予定。

形態素解析

例文テスト

コスト更新

地名語登録

抽出失敗！

JSON形式で返答 / CMS
のモジュールとして利用

地名語抽出

地名語解決

メタデータ付与

GeoNLP+サーバ

2012/11/16

ヒューマンプローブシンポジウム

26

マスメディアの分析

東日本大震災ニュース 2011年3月11日～ 267,000件

2012年 (2339)													
自国新聞	記事数	1	2	3	4	5	6	7	8	9	10	11	12
201221	1	発生	地震予	中心	フィリピン	東	東	東	東	東	東	東	東
201220	2	小笠原諸島	日本の南	南	海上	伊豆諸島	北上	南シナ海	西	東	小笠原	大	東
201219	24	発生	海上	小笠原諸島	北上	伊豆諸島	高	東	東北	東	小笠原	見込み	東
201218	46	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201217	592	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201216	251	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201215	303	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201214	74	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201213	12	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201212	24	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201211	95	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201210	75	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201209	59	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201208	2	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201207	51	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201206	2	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201205	81	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201204	563	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201203	46	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201202	9	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東
201201	1	発生	海上	小笠原諸島	北上	伊豆諸島	小笠原諸島	北上	東	東	小笠原	見込み	東

21	2011-03-11	1382	東日本大震災	日本	被災者	支援	被災地	仏	高松	義援金	放射能汚染	計画停電
20	2011-03-30	1463	被災者	東日本大震災	支援	被災地	被災者	被災地	被災者	会長	計画停電	義援金
19	2011-03-29	1486	被災者	日本	リレー	東日本大震災	被災者	被災地	被災者	代表	計画停電	義援金
18	2011-03-28	1415	被災地	被災者	日大	チャリティ	被災地	被災者	被災地	被災者	被災地	被災者
17	2011-03-27	780	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
16	2011-03-26	916	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
15	2011-03-25	1563	計画停電	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
14	2011-03-24	1480	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
13	2011-03-23	1586	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
12	2011-03-22	1671	被災者	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
11	2011-03-21	1637	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
10	2011-03-20	982	被災者	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
9	2011-03-19	1277	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
8	2011-03-18	1919	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
7	2011-03-17	1680	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
6	2011-03-16	1970	被災者	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
5	2011-03-15	1596	計画停電	被災者	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者
4	2011-03-14	1355	被災者	計画停電	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者
3	2011-03-13	875	被災者	被災者	東日本大震災	被災者	被災者	被災者	被災者	被災者	被災者	被災者
2	2011-03-12	1141	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者
1	2011-03-11	916	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者	被災者

台風ニュース 2003年～ 23,400件

ヤフーニュースで、指定カテゴリ、指定キーワードで自動収集

2012/11/16

ヒューマンプローブシンポジウム

27

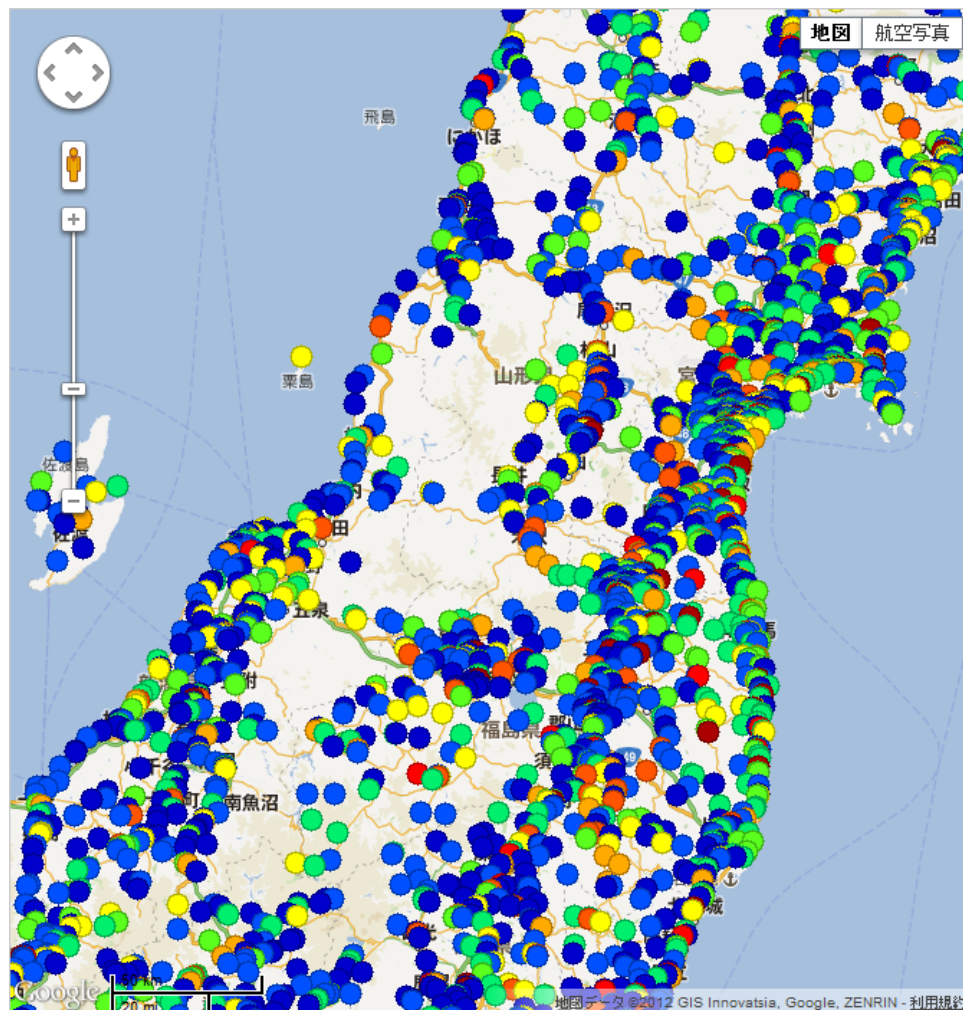
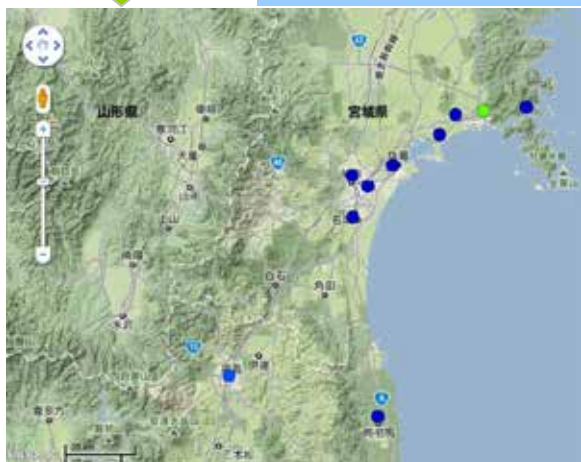
東日本大震災ニュース分析

<http://agora.ex.nii.ac.jp/earthquake/201103-eastjapan/mass-media/>

宮城県などによると、**気仙沼市**や**多賀城市**などで大規模な火災が発生。**気仙沼市**は津波で市街地の3分の1が水没し、**気仙沼湾**に浮かぶ**大島**の4集落が壊滅状態となった。**女川町**もほぼ壊滅という。**岩手県**では**陸前高田市**がほぼ壊滅し、**山田町**や**宮古市**の一部が水没。**福島県**では**南相馬市**の約1800世帯が壊滅状態という。宮城県警は**東松島市**の**JR野蒜駅**付近で、この脱線した列車から乗客ら9人、別の列車からも11人を救出した。

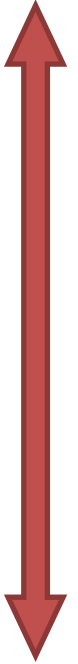


ジオタギング



マスメディアの偏り

市区町村	記事数	人的被害	市区町村	記事数	人的被害
福島	16365	3	宮古	2725	527
仙台	9969	829	大船渡	2690	424
石巻	8424	3735	大槌	2287	1282
気仙沼	4707	1356	名取	2111	966
いわき	4507	347	相馬	1961	458
南相馬	4317	638	女川	1783	915
陸前高田	4131	1795	東松島	1472	1105
南三陸	3039	845	山田	1355	757
郡山	2906	1	山元	939	690
釜石	2864	1046	亘理	846	269



人的被害：総務省消防庁第145報（平成24年3月13日12時現在）による。

記事数：「東日本大震災ニュース分析」 2012年7月下旬現在

2012/11/16

ヒューマンプローブシンポジウム

東日本大震災 ビッグデータワークショップ

<https://sites.google.com/site/prj311/>

- 3月11日から1週間の朝日新聞記事（提供：株式会社朝日新聞社）
- Google Trends（提供：グーグル株式会社）
- 東日本大震災直後のテレビ放送テキスト要約データ（提供：JCC株式会社）
- 3月11日から1週間のツイート（提供：Twitter Japan 株式会社）
- NHK総合テレビ大震災発災直後から24時間の放送音声書き起こし及び頻出ワードランキング（提供：日本放送協会）
- Honda インターナビ通行実績マップデータ（提供：本田技研工業株式会社）
- レスキューナウの鉄道運行情報／緊急情報／被害状況のまとめ情報（提供：株式会社レスキューナウ）
- 混雑統計データ（提供：株式会社ゼンリンデータコム）

- 東日本大震災に関する多種多様なデータを参加者に提供。
- 「次の災害に備える」ためのサービス検討することが目的。
- データの量・種類とも画期的。ただし最長2013年1月まで。

東日本大震災ツイート分析

<http://digital-typhoon.tv/prj311/>

Q. 何が難しいの？ A. 地名の表記が曖昧！

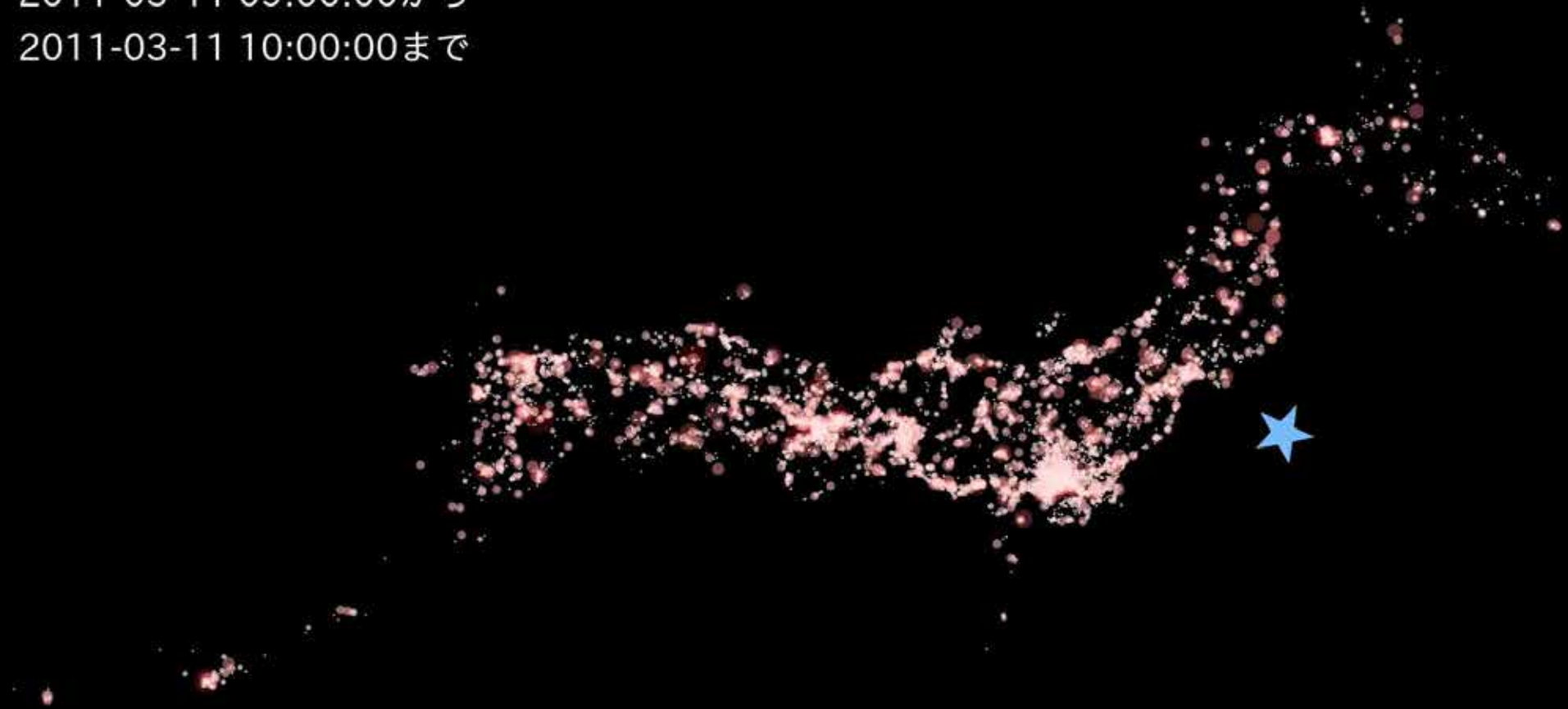
「第一小に避難した」 → 全国に「第一小学校」は269校もある！

全ツイート 179,286,297 → 地名語含むツイート
21,398,123 (12%) を抽出 (参考：ジオタグつき
0.15%)。地名語出現数としては32,956,427。

地名語辞書：約62万（国土交通省、SaveMLAKなど）＋住所のアドレスマッチング。

処理速度：4万ツイート／1コア・時 → 112コアで約40時間かけて処理。

東日本大震災ツイートマップ
2011-03-11 09:00:00から
2011-03-11 10:00:00まで



東日本大震災ビッグデータワークショップ：Twitter Japanのデータを利用

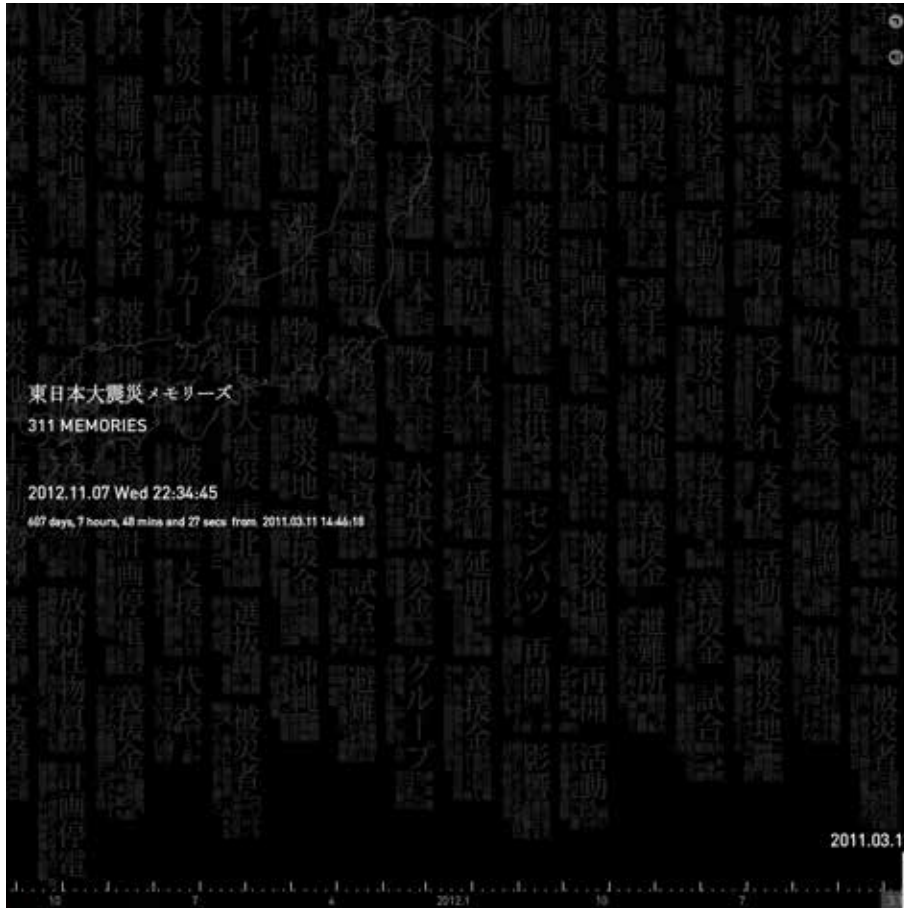
2012/11/16

ヒューマンプローブシンポジウム

32

情報のプッシュ化

<http://agora.ex.nii.ac.jp/earthquake/201103-eastjapan/311memories/>



- **311メモリーズ**：
静かに動く年表。
- ユーザが受動的に
視聴するモードで
は勝手に動く。
- 見ることに集中す
る余裕が生まれ思
索モードに入る。

情報のテレビ化

<http://digital-typhoon.tv/311tv/>



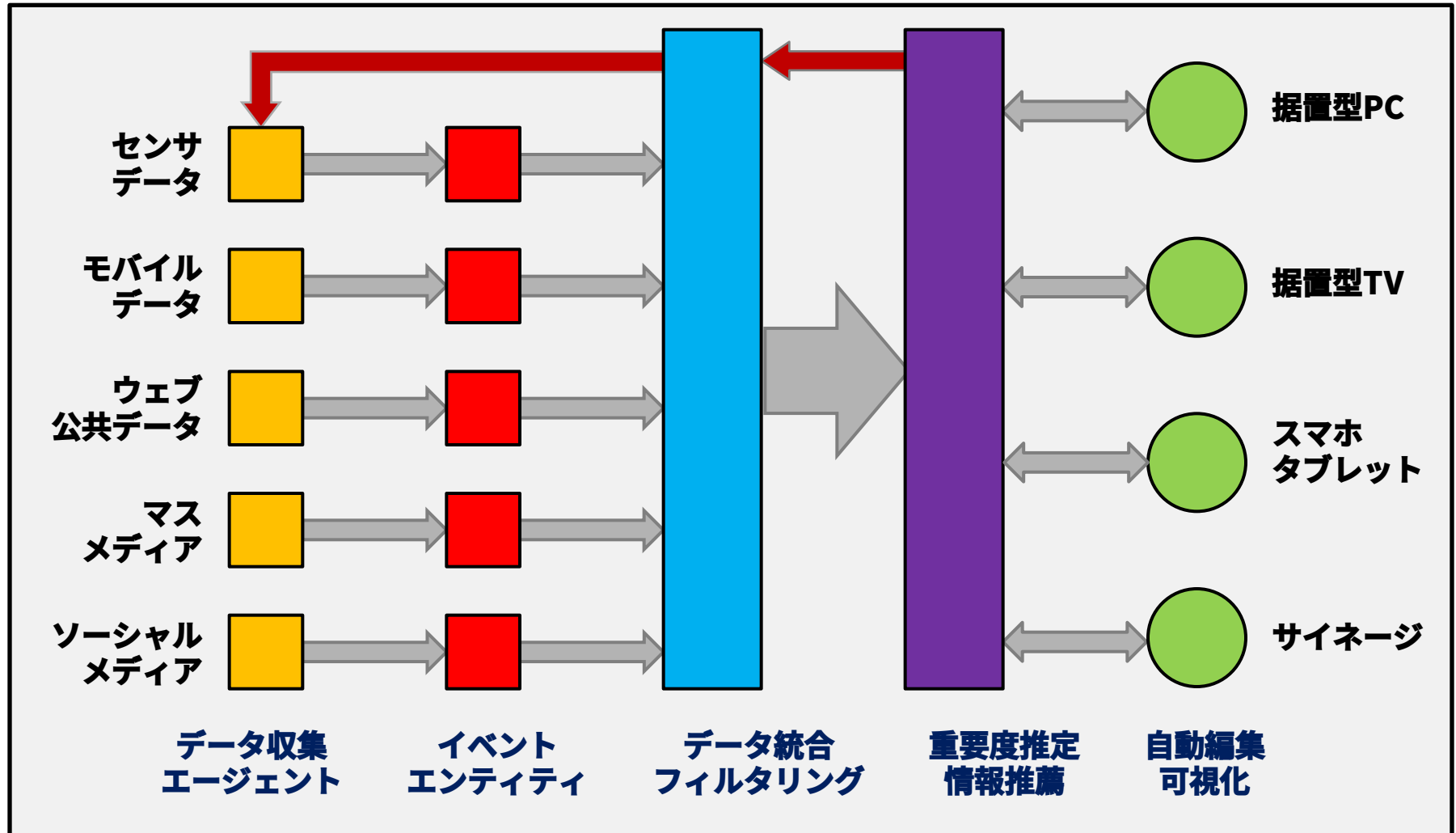
東日本大震災ビッグデータワークショップ：Twitter Japan、朝日新聞、NHK、レス
キューナウ、日本気象協会、ゼンリンデータコム of データを利用

2012/11/16

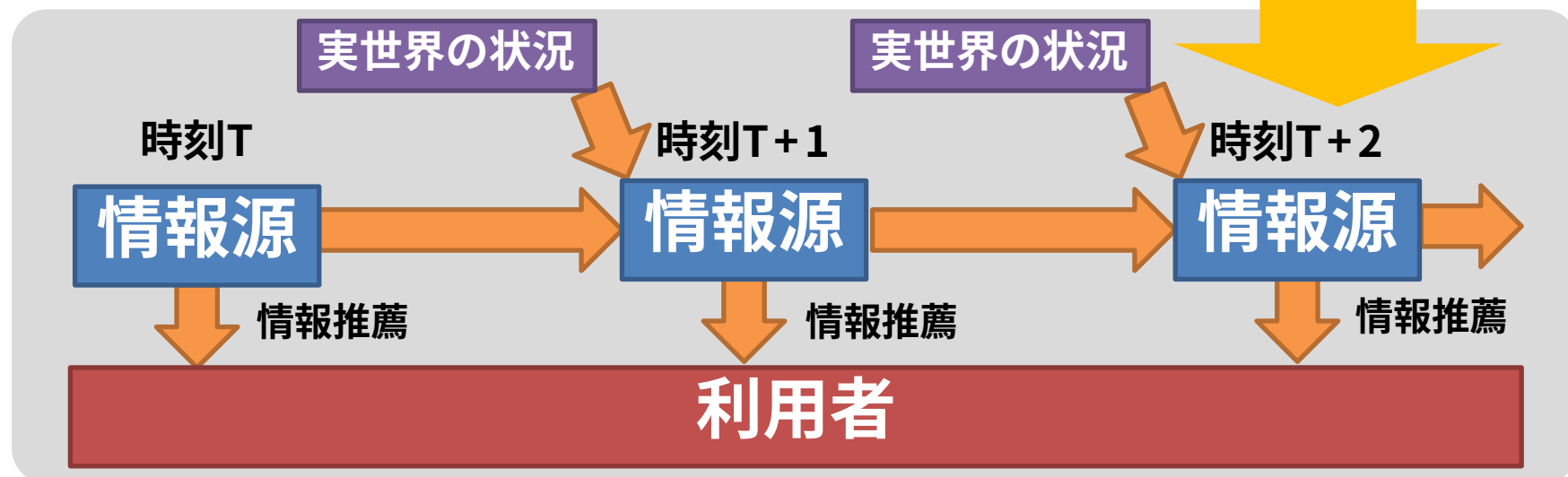
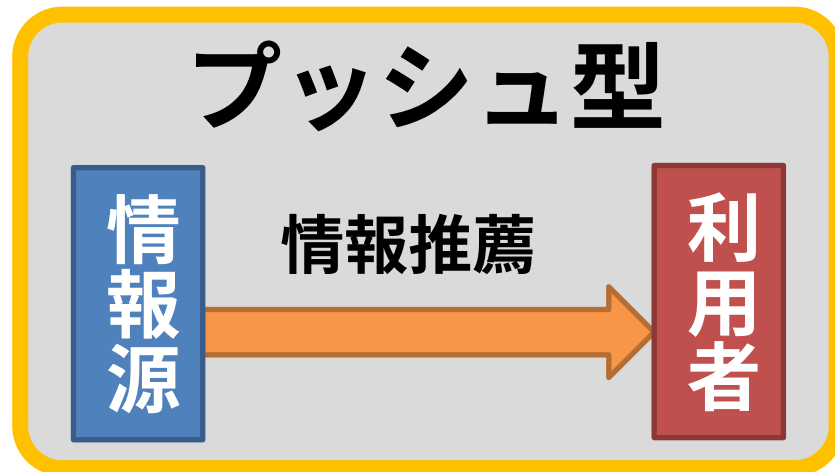
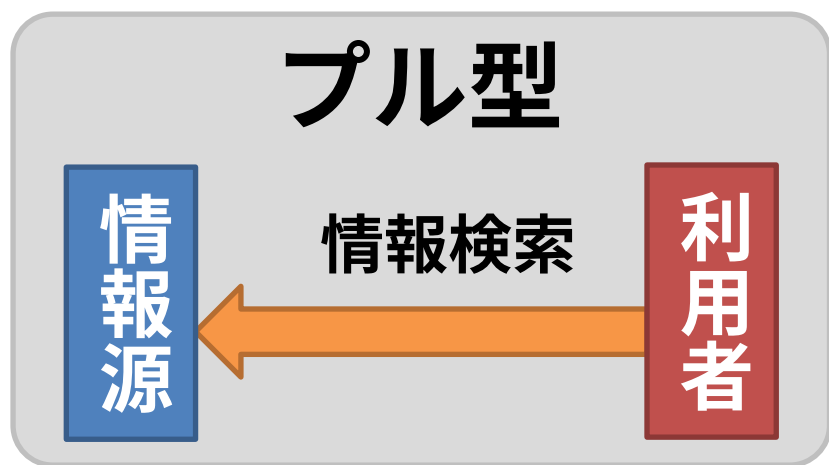
ヒューマンプローブシンポジウム

34

クライシスメディアのデザイン



ベイジアンテレビ



未来社会のメディア

- データの公開とマシンアクセスIBILITY
- データの共有とマッシュアップ
- 複数の情報源の比較
- データや事実を裏付けとした意思決定
- 現在データと過去アーカイブの連続性

ビッグデータ時代の
データ駆動型メディアへ

ヒューマンプロローブの役割

- **HaaV**: センサの移動性を活用した動的な探索（ホットスポットの発見など）。
- **HaaS**: 既存のセンサでは観測しにくい高次情報の取得、心に訴える主観的な情報、将来的には音声も活用？
- **既存のセンサ網の劣化バージョン**にならないための付加価値が必要。
- **次のクライシスに今から備えましょう！**

おわりに

- デジタル台風
 - <http://agora.ex.nii.ac.jp/digital-typhoon/>
- ふってきったー
 - <http://agora.ex.nii.ac.jp/futtekitter/>
- GeoNLP
 - <http://agora.ex.nii.ac.jp/GeoNLP/>
- 個人ページ
 - [http://agora.ex.nii.ac.jp/~ kitamoto/](http://agora.ex.nii.ac.jp/~kitamoto/)

JSTさがけ「知の創生と情報社会」などの支援を受けました。
東日本大震災ビッグデータワークショップのデータ（Twitter Japan、朝日新聞、NHK、レスキューナウ、日本気象協会）のデータを利用しました。