

2018年7月2日

LODチャレンジ2018ミートアップ

@国立情報学研究所

Tweet2LinkDataで始める 参加型オープンデータ

合同会社 緑 I T 事務所

代表 小池 隆

Tweet2LinkDataの発想

- ▶ Twitterでは、データとして価値のあるツイートもすぐにタイムラインの彼方に消えてしまい、勿体ない



- ▶ 参加型オープンデータを始めるのは大変
 - ▶ データ登録サイト／アプリの作成
 - ▶ データ入力が面倒（特に、位置情報）



- ▶ Twitterで簡単にオープンデータ作成

Tweet2LinkDataとは

- ▶ ツイートを用いた、カジュアルな
参加型オープンデータ収集・公開基盤
 - ▶ LODチャレンジ2016基盤技術部門最優秀賞

特定のハッシュタグ、写真、
緯度経度付きのツイートを収集



(手動)

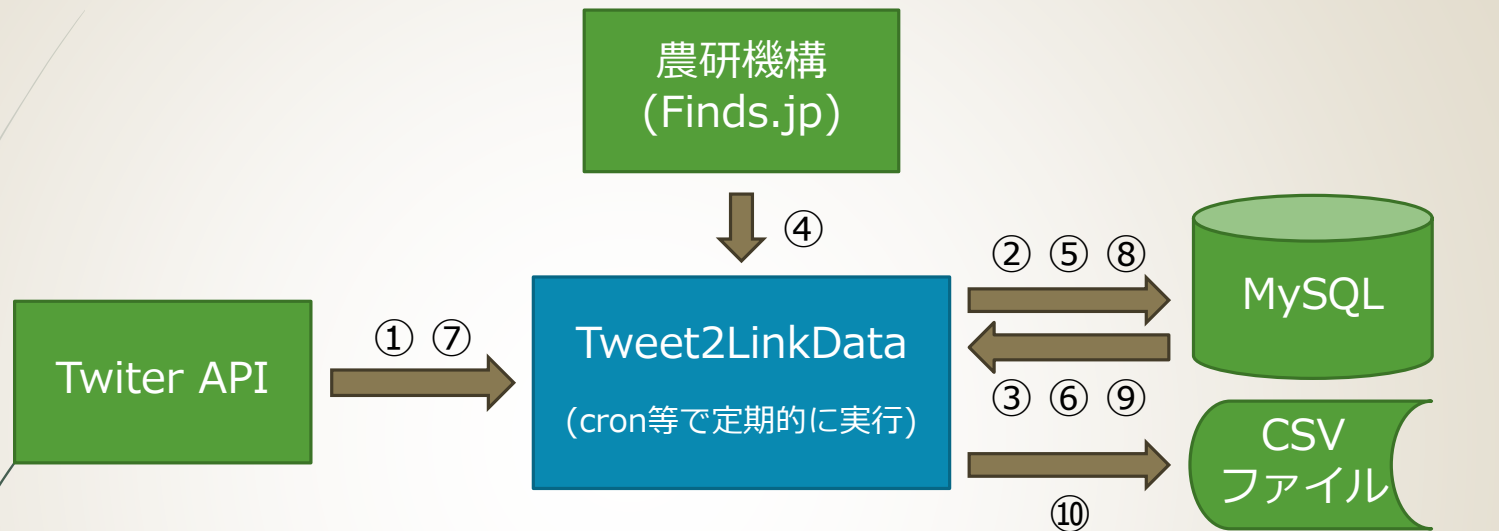
LinkData.orgで公開



(自動)

マップで公開

Tweet2LinkDataの動作



- ① ツイートを取得
- ② 新規ツイートをINSERT
- ③ 住所のないレコードを取得
- ④ ツイートの緯度経度から住所を取得
- ⑤ 住所をUPDATE
- ⑥ 埋め込み用HTMLのないレコードを取得
- ⑦ 埋め込み用HTMLを取得
- ⑧ 埋め込み用HTMLをUPDATE
- ⑨ 全レコードを取得
- ⑩ LinkData用とマップ用のCSVファイルを出力

- LinkData用
- マップ用

動作環境

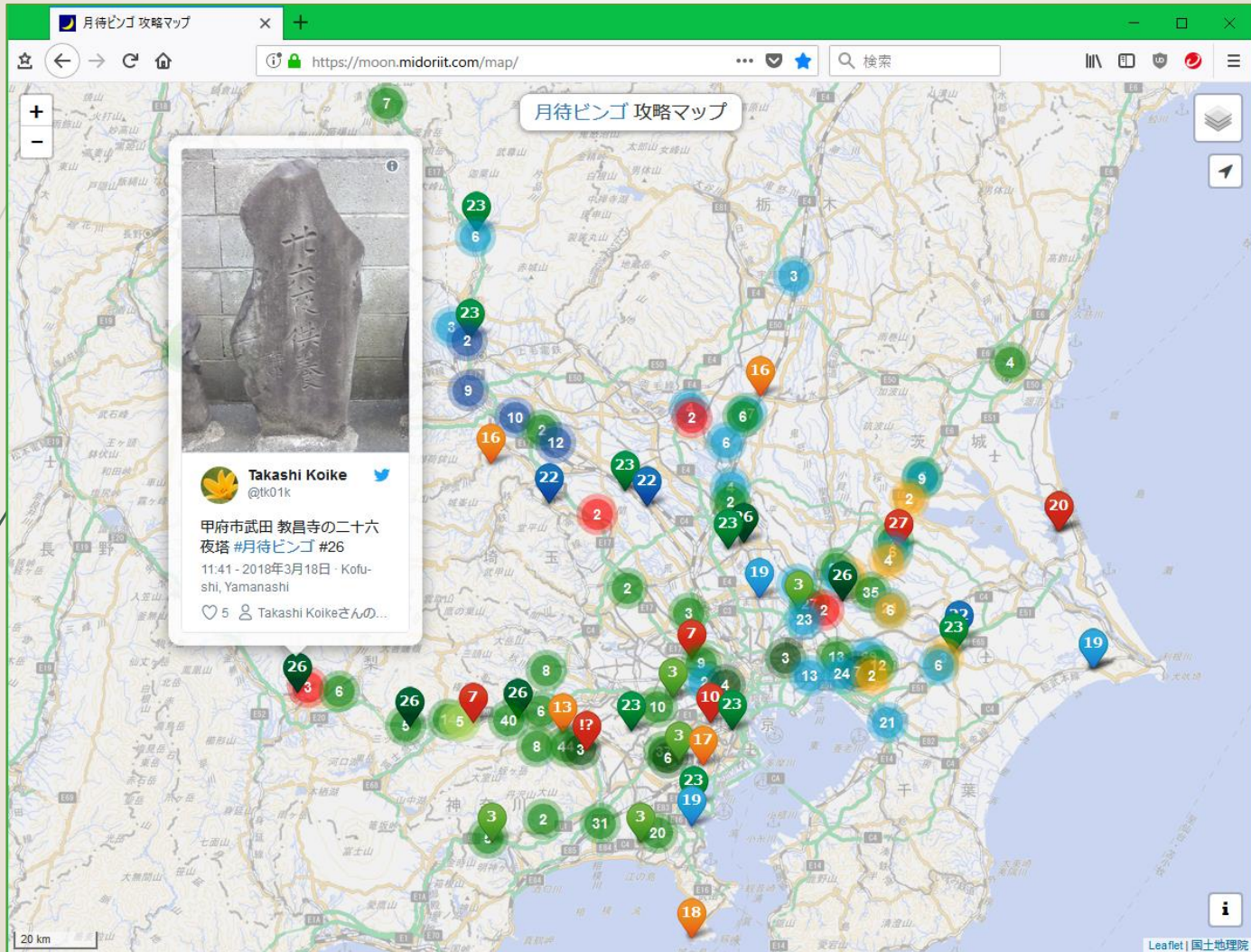
- Webサーバ
- PHP
- MySQL

利用例①：月待ビンゴ

➡ 月待塔をツイートして集める



マップによる公開



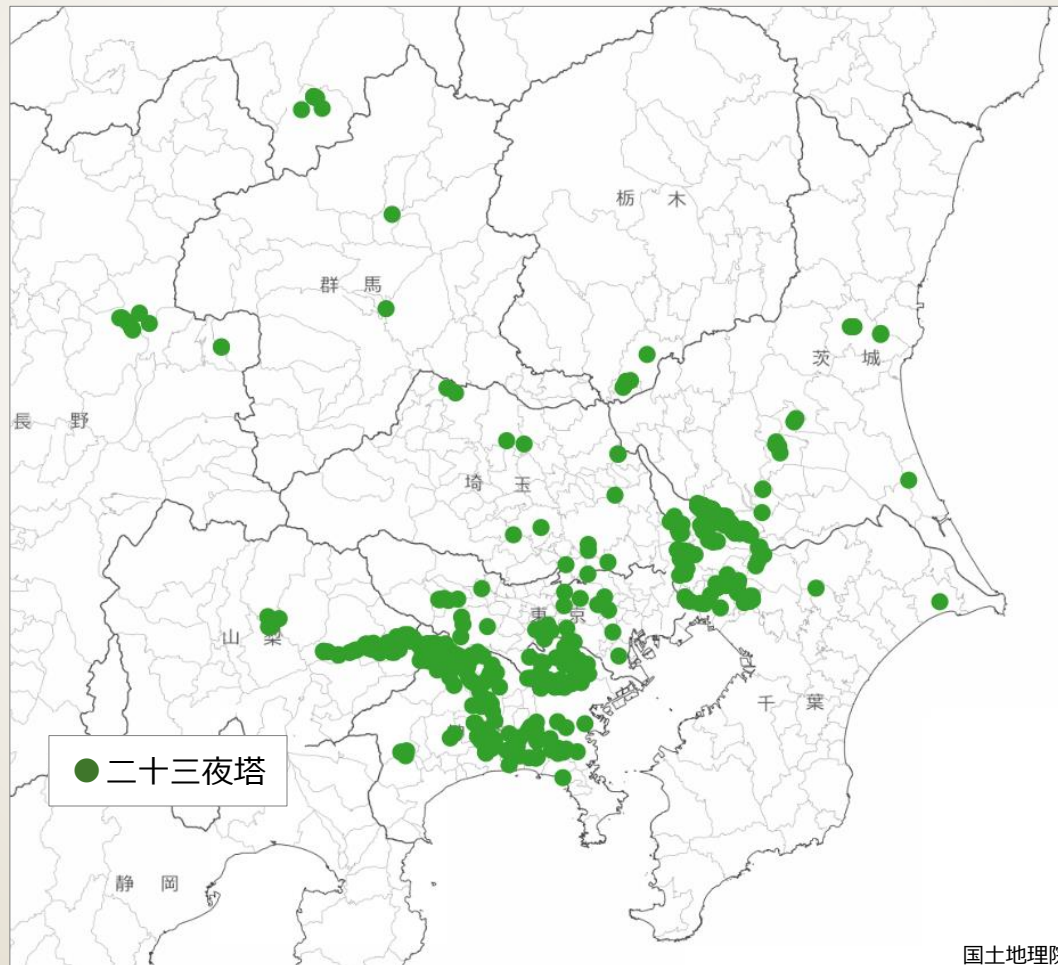
LinkData.orgで公開

- ➡ 870件（2018/6/30版）
- ➡ 塔の種類、紀年銘もデータ化

LinkData.orgのウェブページスクリーンショット。ブラウザのアドレスバーには linkdata.org/work/rdf1s4819i と表示されています。ページ上部には「LinkData」のロゴと「オープンデータを加工して共有しよう テーブルデータの変換と公開をサポートするサイト」というキャッチフレーズがあります。ナビゲーションメニューには「HOME」「チュートリアル」「テーブルデータ作成」「データのアップロード」「データセット一覧」があります。メインコンテンツには「データセットの情報」セクションがあり、「共同編集の招待を送る」ボタンと、神奈川県、長野県、新潟県、茨城県、栃木県、群馬県、埼玉県、千葉県、東京都、山梨県などの地域フィルターがあります。中央には「月待塔」というデータセットの紹介があり、写真、ユーザー名「小池 隆 合同会社 緑 I T 事務所」、および「モバイルビューを開く」ボタンが示されています。右側には「6 評価指数」という評価ボックスと「いいね!」ボタンがあります。下部には、日本各地にある月待塔（民間信仰により造立された石仏・石塔の一種で、十五夜、十九夜、二十三夜のように、月待行事を行う日付によって分類される）を使用したゲーム「月待ピンゴ」(https://moon.midoriit.com) 参加者のツイートを用いて作成した月待塔の種類と位置情報のデータセットです。Tweet2LinkData (http://idea.linkdata.org/idea/idea1s2075i) を利用して自動的に収集・加工しています。

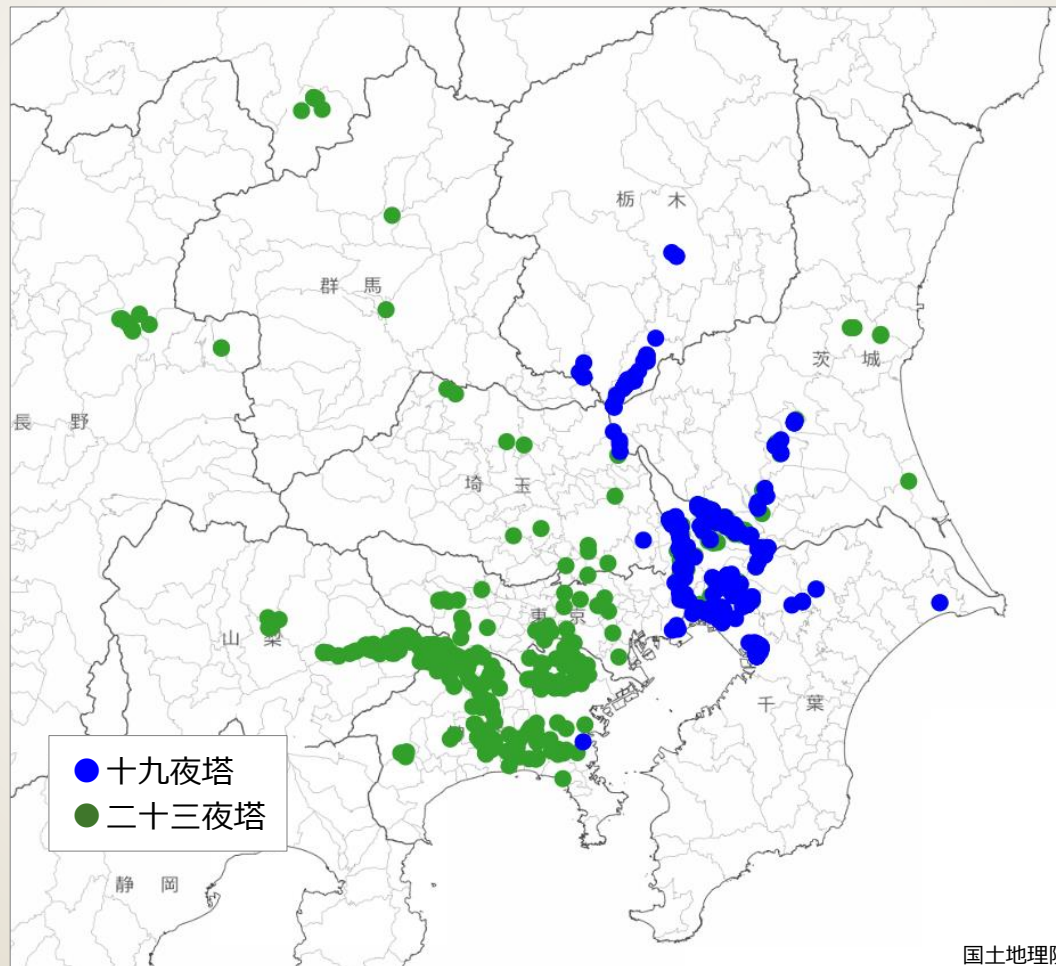
月待塔の種類別分布の可視化

➡ 二十三夜



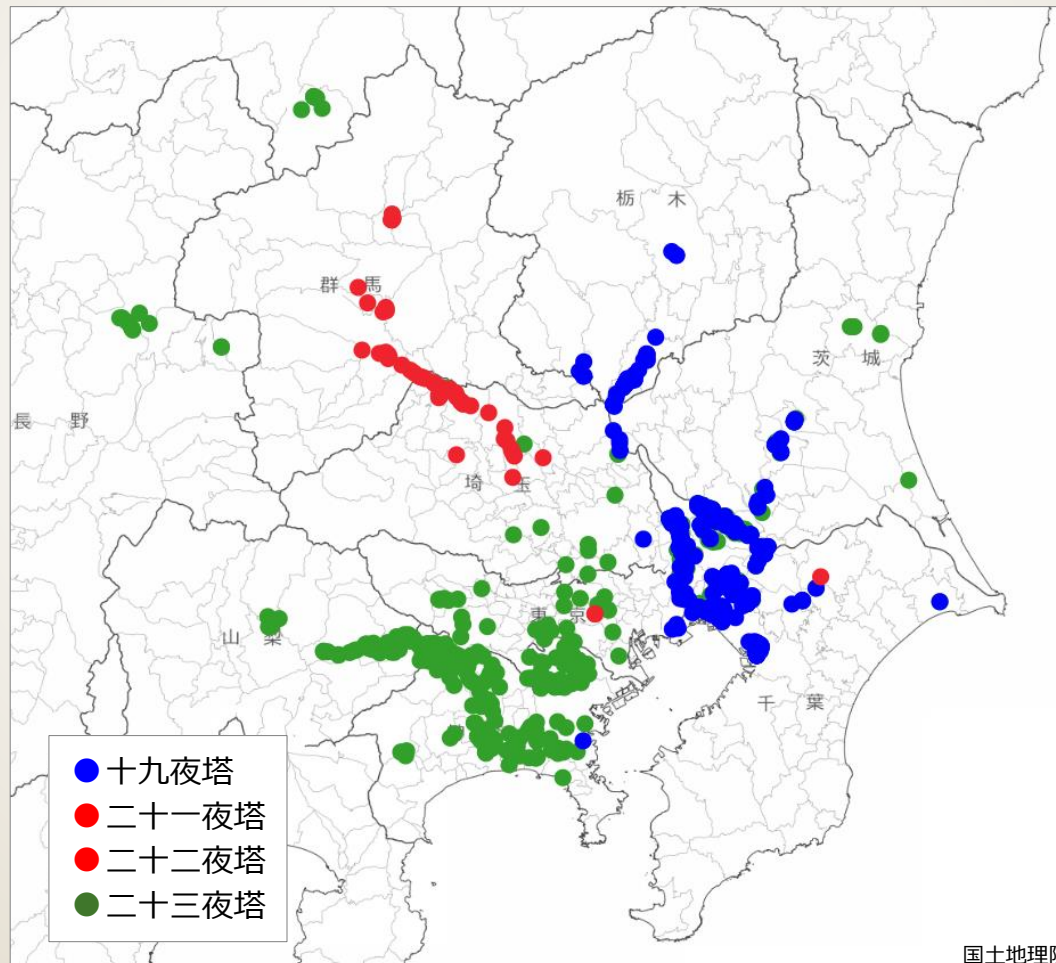
月待塔の種類別分布の可視化

➡ 十九夜、二十三夜



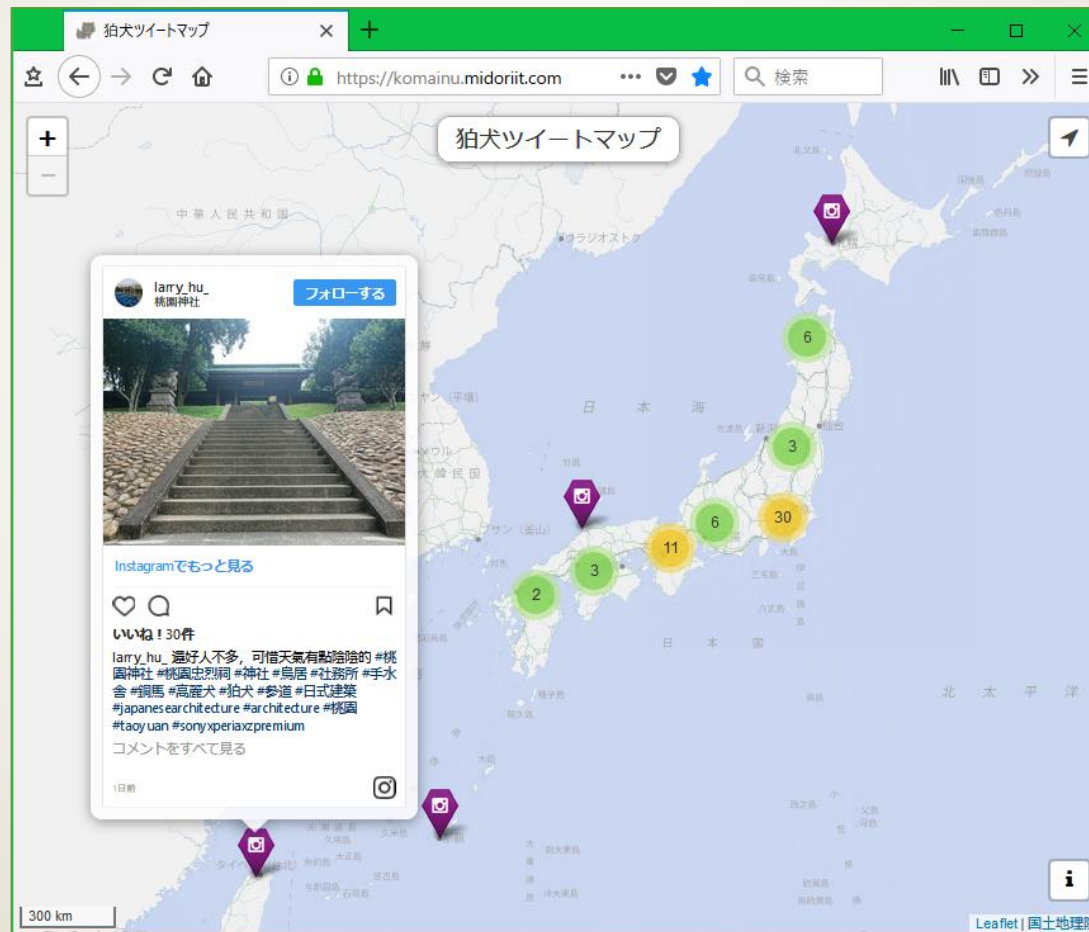
月待塔の種類別分布の可視化

➡ 十九夜、二十一夜、二十二夜、二十三夜



利用例②：#狛犬 ツイート収集

➡ Instagramに対応（Twitter同時投稿の場合）



オープンデータ化以外の活用法

- ➡ データ分析のための収集
 - ➡ 著作権法47条の7
情報解析のための複製等
- ➡ **#狛犬** ツイートの収集データ
(2018/1/1~6/30)
 - ➡ 1,607件のツイートのテキスト
 - ➡ テキストマイニング?
 - ➡ 2,022枚の狛犬 (が写っている
かもしれない) 画像
 - ➡ 機械学習で狛犬検出!?

参加型オープンデータの
収集・公開に
Tweet2LinkDataを
ぜひ、ご活用下さい

github.com/midoriit/Tweet2LinkData