POT抽出について

Notebook: CS

Created: 2018/02/22 11:07 **Updated:** 2018/02/22 12:42

Author: Yki

POI抽出について

【クエリ内容の概要】

Google Places API Place Search

・一つの (lat, lon) Place Search クエリに付き最大60件のPOIレコードを返す(1 ページ・API call/20件のPOIレコード, 60件=3 API calls)

- ・一つの (lat, lon) 地点からradius=50,000m以内までの近所のPOIをクエリできる
- ・POIの詳細情報を見るのに追加で 1 API call が必要
- ・取れる詳細情報:POIの画像、座標、カテゴリー、レビューの内容、評価、営業時間
- · API利用率上限:1日/1,000 API calls

LiveJapan

- ・取れる情報:POIの画像、座標、カテゴリーとサブカテゴリー、営業時間
- ・GP APIより登録POI数が少ないが、観光客にRelevanceが高いPOIが多いと思われる
- ・地図はZenrinの独自サービスを使っているようで、POIのGoogle IDの記載がない(座標・POI名称でマッチングする必要ある)
- ・webscrapingで取ったため、そのまま使えず、Google Places APIなどとのリンクが必要

【POI抽出方法の2パターン】

- 1) Google Places APIのみからPOIを抽出:
- -都営バスの各バス停の座標を中心に1,000mの検索radiusの中からPOIを抽出する -使える結果の数を最大化するために評価点数4以上等、という条件を支持

所有時間

-Food, Tour, Shopの各カテゴリーで最初の20~60件のPOIを抽出の場合: (3~9 api calls x バス亭の数のPlace Searchクエリ)x 2回 (各POIの詳細情報をクエリ)→ バス停100件の場合、最低で約 6.000 API calls/ 所有時間 6日間~

長所:

- -より多くのPOIレコードを抽出できる
- -評価でフィルターしたPOI結果を抽出できる
- -都営バスのバス亭の位置にPOI検索を絞ることができる

短所:

-Live Japanよりカテゴリーやサブカテゴリーの情報が少ないため、お店の種類などで分ける ことが出来なくなる

(例:レストランカテゴリーの分類ツリー;基本的にサブカテゴリーの情報無し)

- 観光客にrelevantではないPOIも含まれる
- POIによってWebsite情報がない可能性あり
- -所有時間が多少長い
- 2) Live Japanからscrapingした2,400件のPOIとGoogle Placesから名称・座標マッチング:
- -各LJのPOIの座標に対して、Google Places APIでカテゴリーと検索radiusを50mに絞り、付近POIをクエリする
- -クエリー結果をLJでのPOIレコードと名称・座標でマッチングし、同じPOIを抽出する
- -マッチングできた場合、特定したPOIに対して詳細情報を改めてクエリ

所有時間

- (2,400件のLJ POIの座標をPlace Searchでクエリ)x 2回(各LJ POIの詳細情報をクエリ)→ (マッチング出来なかった時の追加API callを配慮し)約 5,000 API calls/ 所有時間5日間 ~

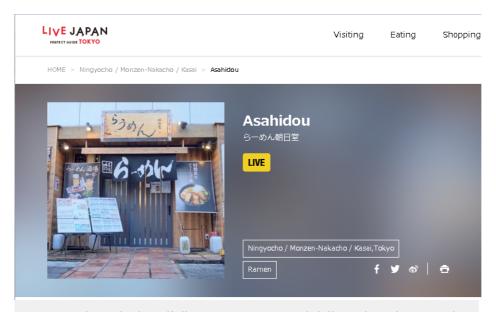
長所:

- -観光客にrelevantなPOIのみに絞れる
- -各POIに対してLive Japanの中でページがある → UIの統一性
- -所有時間が多少短い

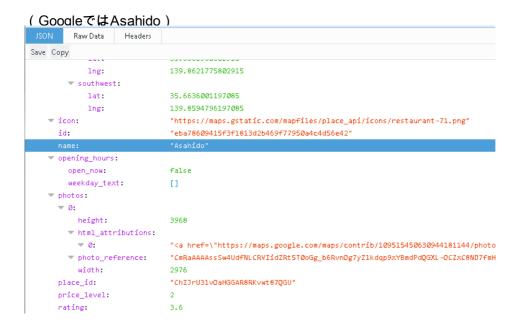
短所:

- -位置でフィルターしていないため都営バス亭と離れている対象外レコードもいっぱい取って しまう
- -バス亭との距離でフィルターして上記を外した後、対象POIレコードが非常に少なくなる可能性
- -GoogleでのPOI名称や住所の表記と完全一致しないため、マッチングを完全自動化できない 可能性あり

(LJではAsahidou)



Second Tsuboi Building 1F,5-6-11, Nishikasai, Edogawa-ku, Tokyo



vicinity: "5 Chome-6-11 Nishikasai, Edogawa"