



geocoder.ja

「Meet the Orkneys & 交流会 @横浜」

2006. 11. 1

(株)ジオセンス

小林一英

自己紹介

- (株)アルプス社に30数年
- 商品開発、システム開発を担当
- 森さんとベストセラーを連発

1980-1998: 地図というテーマが良かった

- 失敗作: ジオアトラス 98.10
GoogleMapsの施設検索表示と同じ機能
ジオコーダそのもの
- 現在は受託開発
地図、GIS、ハード開発、何でも

ジオコードとは

- **住所を緯度経度に変換**

アドレスマッチングとも言う

横浜市中区海岸通1-2 → 35度6分56秒 139度38分29秒

- **Geocode** : 地理的なコード。

- 緯度経度 コード?
- メッシュコード
- 郵便番号
- 固定電話番号
- ロカポ (LocaPoint)
- マップコード (3m)
- ユビキタスの場所コード (ucode: 3cm)
- Georef
- Natural Area Code

ジオコーダとは

● 基盤技術

○道路データ(写真) + (ジオコーダ + 住所リスト) = 地図

(例) Google Geocoder API + 電話帳データ = 住宅地図

○完全なジオコーダ + 完全な経路探索 + 1m精度の位置特定技術

↓
地図不要、迷わない > コストダウン、エコ

● オブジェクト特定技術

住所そのものでなく、そこに存在する建物(オブジェクト)が関心事。
知りたいのはオブジェクトの位置と到達方法。

ピンポイントマッチング >> 部屋まで出来れば殆ど困らない

ジオコーディング用データ

- 住所と緯度経度の対応付け

- 基本は不動産登記法第14条の「地図」

1/250～1/2,500 精度20cm(市街地)、75cm(村落)

- 「地図」が無い所は「地図に準ずる図面」=公図

既存の2,500図面に嵌め込んで緯度経度を取得



大変な作業 >> 地図会社の存在価値

都市再生街区基本調査(H16-18)のデータに期待

- 住居表示地区は「住居表示新旧対照案内図」を使用

ジオコーディング用データ

- 入手可能データ

- 無料のデータ

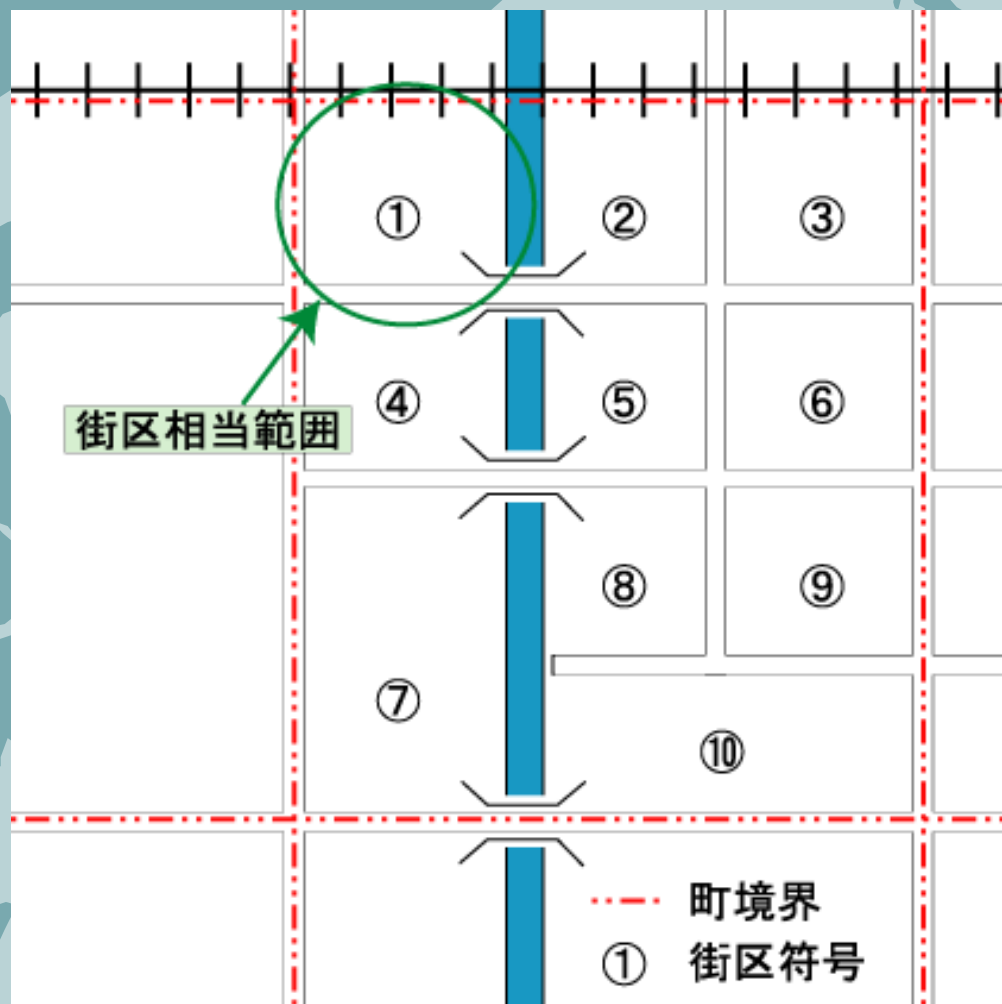
- 国土交通省: 街区レベル位置参照情報

整備エリア: 都市計画区域内のみ

- 有料(数100万レベル?)

- インクリメントP(株)
- (株)アルプス社: 番地号(地番)ポイントデータ
- 昭文社: MAPPLEアドレスデータ

街区レベル位置参照情報



- ・ 国土交通省が整備
ホームページからDL可能
- ・ 都市計画区域内のみ
- ・ 街区の中心の座標
- ・ 毎年更新
- ・ 利用は出典を明記 すれば
無料

既存のジオコーダ

| | エンジン | データ |
|------------|-----------|-------------|
| GoogleMaps | 独自 | ゼンリン |
| YahooMaps | 独自 | アルプス |
| Mapion | 独自 | アルプス |
| Mapfan | 独自 | インクリメントP |
| ESRI | 独自 | インクリメントP |
| MapInfo | 独自 | アルプス |
| CSIS | 独自 | 街区レベル位置参照情報 |
| マップログジオコーダ | 独自 | 街区レベル位置参照情報 |
| マップクエスト | 独自 | 街区レベル位置参照情報 |
| ジャスミンソフト | 独自 (Java) | 街区レベル位置参照情報 |

この表はネット検索で調べたもので、一部であり、単なる推測です。

Geocoder.jaの特徴

- オープンソース (LGPL) + オープンデータ
= 無料 & カスタマイズ自由

データ: 郵便番号データ、街区レベル位置参照情報

- 実用性

商用データ (IPC) もサポート。多様な表記方法に対応。

- 高速

号までのマッチング 30,000件/sec (3GHz, 1GB)

専用DB構造 (オンメモリ、ハッシュ) により高速化

高速化の技法

- **オンメモリ** --- 高速化の鍵

街区レベル: 300MB

インクリメントP: 500MB

- **ハッシュ**使用。If文を使わない。

- ・ 文字列を一定の範囲内の整数に変換
- ・ 地名の頭2文字を4バイト整数と見なし、サイズの余りを取る。文字数は住所のトークンより短い事が必要。
- ・ 県名、郡名、市町村名は一つのハッシュテーブル
+ 町大字名ハッシュテーブル
- ・ 丁目、字、番地、号はシーケンシャル、バイナリサーチ

実装方法に関して

- 郵便番号＋街区レベルデータにした理由
都市計画区域外のサポート。
 - ともかく町名のリストがあるので、郵便番号データ
 - 都市計画区域外の緯度経度はどうする？
wiki的に皆で登録？

プログラムの構成

アプリケーションプログラム

ジオコーダAPI

ジオコーダライブラリ

住所DB

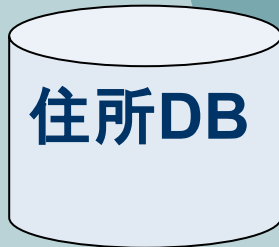
住所データベース作成プログラム

どちらか一方

住所基本データ
郵便番号+街区レベル位置
参照情報

住所基本データ
インクリメントP(株)

住所DBの構成



1. 行政区域レコードファイル
 2. 行政区域ハッシュテーブルファイル
 3. 町大字レコードファイル
 4. 町大字ハッシュテーブルファイル
 5. 丁目字レコードファイル
 6. 地番街区符号レコードファイル
 7. 枝番住居番号レコードファイル
 8. 地番文字列レコードファイル
 9. 複数郵便番号ファイル
 10. 郵便番号インデックスファイル
- の10のファイルから構成

プログラムの概要

- データベース作成

- ・ Windowsで動く6本のプログラム
- ・ ソースはVC++

- ライブラリ

- ・ Linux : gcjlib.so, gcjlib.a
- Windows : gcjlib.dll
- ・ ソースはC
- ・ 地名文字コード: シフトJIS、UTF8、EUCJPに対応
DB内部はシフトJIS

ライブラリAPI

// データベースのロード

```
path = "/data/address_data"; // 住所データベースの格納ディレクトリ  
loadLevel = 4; // 地番街区符号のデータまでをロード  
retCode = gcjDbLoad( path, loadLevel );
```

// 住所を住所コードに変換

```
ADR_CODE adrCode;  
address = "東京都千代田区丸の内一丁目2番3号"  
retCode = gcjAdrStr2Code( address, &adrCode );
```

// 住所コードを緯度経度に変換

```
retCode = gcjAdrCode2Point( &adrCode, &latitude, &longitude);
```

// 住所コードを住所文字列に変換: 正規化

```
retCode = gcjAdrCode2Str( &adrCode, kanji, 256, NULL, 0 );
```

マッチングしない理由

● 表記の違い

1. 字体

- ・ヶ,ヶ,が,ガ ノ,の,之 など
- ・淵,淵 曾,曾 籠,籠 岡,丘 富,富 など
- ・二三、二十三,23,23 など

2. 略し方

- ・一丁目2番3号 1-2-3
- ・“大字”や“字”のある無し
- ・郡名、大字名、字名のある無し

3. 丁目、地番の区切り符号

丁目,町目,番町,番丁,丁,番,線,地割,号,番地,番地の,街区,区,組

マッチングしない理由

- 新旧の違い

 - データベース、検索町名

- 通称名、京都通り名

- 文字コード

- 実際と住所の違い

- その他色々

 - 東京都三宅支庁

 - 東京都三宅島三宅村伊豆642番地

 - OK: GoogleMaps、Mapfan(伊豆まで)

 - NO: YahooMaps, マピオン、ちず丸、its-moNavi

 - “三宅島”は旧郡名 八丈島青ヶ島村

今後の展開

● マッチング率の向上

マッチングしない住所の収集が鍵。

ジオコードに関するサイトの立ち上げ

● 最強のジオコードを目指す

・2.5万地形図の地名検索 36万件 済み

・1文字からの検索 済み

・京都通り名検索 作成中

通りのベクトルデータ作成。 www.geosense.co.jp

地図上でクリックすると通り名(〇〇通XX上ル)を表示等

・全ての地名、位置を検索できるエンジン 企画中

(例)与謝野蕪村生誕地

表示

検索

ツールボックス

- リンク元
- リンク先の更新状況
- アップロード
- 特別ページ
- 印刷用バージョン
- この版への固定リンク
- この項目を引用

他の言語

- Deutsch
- English
- English
- فارسی
- Nederlands

1.2.2 その他

- 2 地名の研究
- 3 ヨーロッパの地名
- 4 関連項目

地名の成り立ち

[\[編集\]](#)

日本の場合

[\[編集\]](#)

日本の地名の半分ほどは、「山」「川」「田」「海」「森」などのもともとあった自然や、特定の自然物（「犬」「魚」「稲」「木」など）から採られた字を含む。それらの地名は比較的古くに名付けられたものが多い。それら長く残ってきた地名に対して、比較的新しい地名は、人工物（「家」「町」「村」「屋」など）や無形物（文化や言い伝えなど）を由来とするものが多い。また、古くからある地名は**万葉仮名**などが最初に当てられた文字になっているものが多く、後に**難読地名**となっているものが多いし、由来を探るのが大変難しい。

日本の地名に多い字

日本の地名に多い字について、国土地理院の空間データ基盤25000（25,000分の1地形図）の地名に関して調べると次のような結果となる。

町,川,目,丁,田,山,大,野,中,沢,上,原,谷,下,西,東,三,一,小,二,ノ,場,新,島,木,南,本,北,内,寺,高,平,井,神,屋,崎,の,松,地,宮,石,水,四,戸,台,ヶ,日,瀬,長,一,八,五,尾,久,津,前,岩,池,村,和,ン,口,岡,浜,坂,古,倉,城,見,浦,吉,園,丘,郷,市,子,森,根,保,ル,生,桑,滝,江,畑,白,ム,里,金,河,塚,六,十,福,泉,黒,部,天,フ,之 （上位100の多い順）

日本以外の場合

[\[編集\]](#)

中国の場合

[\[編集\]](#)

日本と異なり、**戦乱**などによって集団移動することが多かった**中国**においては、元の出身地を姓にす

KYOTO Street Map - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 移動(Q) ブックマーク(B) ツール(T) ヘルプ(H)

http://www.geosense.co.jp/map/kamap-kyoto/htdocs/ 移動

京都通り名マップ 東西の通りは赤、南北は青で表示してあります。 INFO HELP

Kyoto Street Map 1:100000

Map Info
このページはテスト中のものです。

Keymap

Coords
 m: -16404.16
 y: -113682.58

Extents
 minx: -28380.96
 miny: -114529.24
 maxx: -16227.77
 maxy: -103240.36

Kyoto street list

三条通
四条通

完了

古都にたたずむ新築マンション特集～京都・滋賀・奈良～ / 住宅情報ナビ

京都府京都市中京区寺町通御池下る下本能寺寺前町507、508. 京の中心・洛中。華やかな賑わいを継承しながらもしっとりと佇む「寺町」に光溢れる全戸角住戸で誕生。京都市役所前駅徒歩1分をはじめ4線4駅利用可。...

www.jj-navi.com/house/adv/msn/buysell_kansai_m200510kt.html - 82k - 補足結果 -



今後の展開

- 住所の問題点 = マッチングしない理由

住所を使う目的:

位置特定 \neq 住所



オブジェクト特定



オブジェクト識別: オブジェクトID

オブジェクトIDと緯度経度の関連付け

ucode ?

オープンソース開発について

- **見栄**がいいものを作る
人に見られるから、いいものにしなければという見栄。会社ではともかく動けばいい。
- **コスト**をかけても文句を言われたい。
会社はコストをかけた分の売り上げを要求される。
- 作ったものが残る**安心感**。
会社では販売中止になると、精魂かけたものも無になる。
- **淡い期待**
もしかして、名前が知られるようになるかもしれないという...



オープンソース開発は**創作意欲が沸く構造**があると実感