Displaying People with Old Addresses on a Map

Gang Zhang(✉) and Harumi Murakami

Graduate School for Creative Cities, Osaka City University,
3-3-138, Sugimoto, Sumiyoshi, Osaka 558-8585, Japan
springsimb@gmail.com

Abstract. This paper proposes a method of converting old addresses to current addresses for geocoding, with the aim of displaying on a map people who have such old addresses. Existing geocoding services often fail to handle old addresses since the names of towns, cities, or prefectures can be different from those of current addresses. To solve this geocoding problem, we focus on postal codes, extracting them from Web search result snippets using the query "prefecture name AND important place name AND *postal code*." The frequency of postal codes and the edit distance between the old address and the addresses obtained using the postal codes are used to judge the most suitable postal code and thus the corresponding current address. The effectiveness of the proposed method is evaluated in an experiment using a relative dataset. A prototype system was implemented in which users could display people using their birthdate and birthplace addresses on a map chronologically with an associated history chart.

Keywords: People search · Old address · Postal code · Edit distance · Map display · Information extraction

1 Introduction

In our private lives, it is very important to learn about our ancestors and relatives and organize their information. Marking their positions on a map is a useful and feasible approach. The aim of this research is to develop a map interface to access people from the past to the present. Geocoding services are available to display people using their addresses. We want to obtain addresses from documents like census registers or other relevant documents. However, existing geocoding services often fail to obtain location information from old addresses. Most of the addresses contained in documents like family registers follow an older addressing system out of use today. It is difficult to convert old addresses to currently used addresses because house numbers, region names, and even basic town names change over time.

To solve the geocoding problems caused by old addresses, we focus on postal codes found on the Web. We assume that useful information for finding current addresses exists on the Web, and postal codes are good clues since they are easy to extract and easy to convert to current addresses using data provided by Japan Post [1].

G. Zuccon et al. (Eds.): AIRS 2015, LNCS 9460, pp. 381–386, 2015.
DOI: 10.1007/978-3-319-28940-3_30

The frequency of postal codes and the edit distance between old addresses and addresses obtained from the postal codes are used to identify the most suitable postal code and thus the corresponding current address. The effectiveness of the proposed method is confirmed in an experiment using a relative dataset that contains addresses from 1905 to 2006. A prototype system was implemented in which users could display people using their birthdate and addresses on a map chronologically with an associated history chart.

The structure of this paper is as follows. In Sect. 2 we explain our method. Experimental results are described in Sect. 3. The prototype is shown in Sect. 4. Related work and future work are described in Sect. 5. The examples presented in this paper were translated from Japanese into English for publication.

2 Approach

2.1 Overview

An overview of the proposed method consists of three stages:

(a) postal code collection: collects postal codes relevant to an old address from the Web; (b) best postal code judgment: calculates scores of postal codes using the frequency of each postal code and the edit distance between the old address and an address associated with the postal code obtained from Japan Post postal code data [1], and then judges the best one; and (c) address refinement: refines the address using Ministry of Land, Infrastructure, Transport and Tourism (MLIT) positioning reference data [2].

Table 1 presents example results from using the proposed method. We arbitrarily set house numbers as "99" in order to protect personal information. In this paper, we discuss example number 1.

Table 1. Example results of proposed method

No	Address in family register	Output address by proposed method
1	99-banchi, 1-chome, Uchiandojimachidori, Minami-ku, Osaka-shi	1-chome, Andojimachi, Chuo-ku, Osaka-shi, Osaka-fu
2	99-banchi, Kitanokofukacho, Kita-ku, Osaka-shi	1-chome, Shibata, Kita-ku, Osaka-shi, Osaka-fu
3	99-ban-yashiki, Takagi, Kokufu-mura, Ashida-gun, Hiroshima-ken	Takagi-cho, Fuchu-shi, Hiroshima-ken

2.2 Extracting Postal Codes

Addresses change with time. We focus on postal codes since they can be easily converted to the current addresses. To search for the postal code of an address, we use Web search engines; in this paper, we use Google. The combination of important place name and prefecture name were assumed to be a good choice for the query sent to the search engines, and thus the postal codes included in the snippets of the results were collected.

Extracting Important Place Name. We assume the most important type of place name that "lives long" is town names. To extract a Japanese town name, we remove the character strings focusing on numerical positions and such positional keywords as aza, mura, cho (town), gun, ku (ward), shi (city), to, do, fu, or ken (prefecture). For example, "Uchiandojimachidori" is extracted as an important place name.

Extracting Prefecture Names from the Web. To obtain a prefecture name, the old address without house numbers (e.g. 99-banchi, 99-ban-yashiki) is used for the query to the search engine. When the query "*prefecture* AND old address without house numbers" is given, our method obtains five snippets of results from Google. Prefecture names are extracted by matching a prefecture dictionary to the snippets and thus obtaining the most frequent prefecture name. For example, when the query "*prefecture* 1-chome Uchiandojimajidori Minami-ku Osaka-shi" is given, the result "Osaka-fu" is obtained.

Extracting Postal Codes from the Web. When the query "prefecture name AND important place name AND *postal code*" is given to Google, our method obtains one result snippet, and it extracts postal codes by using the following regular expression from the snippet: `\d{3}-\d{4}`.

When there is no result, our method repeats the above process for the other four result snippets. When no result is found even after five snippets, our method ends and *none* becomes the answer. For example, when the query "Osaka-fu AND Uchiandojimachidori AND *postal code*" is given, such postal codes as 542-0061, 542-0067 and 539-0000 are obtained from the snippet of the top page.

2.3 Judging Best Postal Code

We assume that the more frequently appearing postal codes are related to the current address, and the more similar character strings are related to the current address. Therefore, we calculate scores of postal codes by using the frequency of postal codes and the edit distance between addresses to judge the most suitable postal code. We use Japan Post postal code data to find an address associated with a postal code.

Edit distance is a way of quantifying how dissimilar two strings are to one another by counting the minimum number of operations required to transform one string into the other. We use Levenshtein distance as the edit distance.

The score is calculated as follows:

$$score = \frac{f}{d} \tag{1}$$

Here, f is the frequency of postal codes and d is the edit distance between an old address string without house numbers and an address associated with the postal code.

We select one postal code having the highest score. For example, the frequency of postal code 542-0061 is 3 and its edit distance is 11 (in Japanese), and thus its score (3/11=0.273) becomes the highest among the collected postal codes. Consequently, the address associated with the postal code 542-0061 is obtained as “Andojimachi, Chuo-ku, Osaka-shi, Osaka-fu.”

2.4 Address Refinement

Most of the current official Japanese town names consist of a “chome” (district of a town) and a “town”. The “chome” information is seldom contained in Japan Post data. Very old addresses do not have the “chome” designation either. We add a “chome” to the address obtained in Sect. 2.3 to complete the official town name: the available chome information is used when the old address includes it; otherwise, “1-chome” is used tentatively. Next, we check MLIT position reference data provided by the Ministry of Land, Infrastructure, Transport and Tourism, which contain current official addresses, and if the address exactly matches an entry, it becomes the final answer; if not, the “chome” is deleted. For example, “1-chome, Andojimachi, Chuo-ku, Osaka-shi, Osaka-fu” becomes the answer.

2.5 Geocoding

We obtain latitude/longitude information of the addresses obtained in Sect. 2.4 by using Google Maps Geocoding API v3 [3].

3 Experiment

3.1 Method

We use a relative dataset created in a previous work [4]. The oldest year with an address (birthplace) was 1905, and the newest year with an address (death place) was 2006. Since this is a dataset for Japanese relatives from the modern era, many addresses are identical. Thirty-two different addresses of birth or death places are used in the experiment.

The proposed method was compared with a method that gives old addresses without house numbers directly to Google Geocoding API v3.

We investigated current addresses for the old addresses without house numbers, in other words, town (including chome) levels at each public office corresponding to the old address. We evaluated the answers (output by the two methods) at prefecture, city, and town (including chome) levels. Here, traditional precision and recall are used for evaluation:

$$precision = \frac{r}{n} \tag{2}$$

$$recall = \frac{r}{c} \tag{3}$$

where r is the number of obtained correct answers, n is the number of answers output by the methods, and c is the total number of addresses in the dataset.

Table 2. Results of experiment

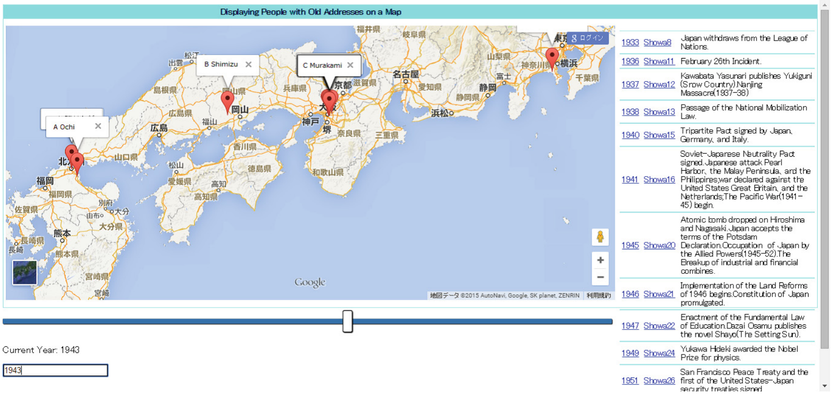
	Proposed method				Google geocoding			
	Precision		Recall		Precision		Recall	
Prefecture	100 %	(31/31)	97 %	(31/32)	96 %	(24/25)	75 %	(24/32)
City	87 %	(27/31)	84 %	(27/32)	80 %	(20/25)	63 %	(20/32)
Town	64 %	(16/25)	64 %	(16/25)	60 %	(9/15)	32 %	(8/25)

3.2 Results

Table 2 shows the results of the experiment. Our method outperformed Google Geocoding API, particularly in recall.

4 Prototype

We have developed a prototype system that enables users to display people on a map using Google Maps JavaScript API v3 [5] through a timeline. Figure 1 shows living people in 1943 on a map with birthplace addresses. This application has the following functions: (a) users can designate the year to display by inputting it or by using a slider tool; and (b) users can browse historical events (extracted from a history book [6]) before and after the year in a history chart, and they can also change the year by selecting a historical event in the chart.



Current Year: 1943

1943

1933	Showa8	Japan withdraws from the League of Nations.
1936	Showa11	February 26th Incident.
1937	Showa12	Kawabata Yasunari publishes Yukiguni (Snow Country).Nanjing Massacre(1937–38)
1938	Showa13	Passage of the National Mobilization Law.
1940	Showa15	Tripartite Pact signed by Japan, Germany, and Italy.
1941	Showa16	Soviet–Japanese Neutrality Pact signed.Japanese attack Pearl Harbor, the Malay Peninsula, and the Philippines;war declared against the United States Great Britain, and the Netherlands;The Pacific War(1941–45) begin.
1945	Showa20	Atomic bomb dropped on Hiroshima and Nagasaki.Japan accepts the terms of the Potsdam Declaration.Occupation of Japan by the Allied Powers(1945–52).The Breakup of industrial and financial combines.
1946	Showa21	Implementation of the Land Reforms of 1946 begins.Constitution of Japan promulgated.
1947	Showa22	Enactment of the Fundamental Law of Education.Dazai Osamu publishes the novel Shayo(The Setting Sun).
1949	Showa24	Yukawa Hideki awarded the Nobel Prize for physics.
1951	Showa26	San Francisco Peace Treaty and the first of the United States–Japan security treaties signed

Fig. 1. Prototype system

5 Related Work and Discussion

FamilySearch [7] enables users to display people on a map based on manual input of their prefecture-level location. Yamamoto et al. [8] proposed techniques for achieving the geographical navigation of historical events described in

Web pages as a Virtual History Tour. They extracted historical event sentences from the Web and identified one representative place name for each event. The place names were mainly at the prefecture or city level. However, we converted addresses in family registers to current addresses and focused on the town level.

Many Web sites provide a facility that converts addresses or landmarks into latitude/longitude, such as geocoding services. In general, however, they cannot handle old addresses. We aim to handle both old and current addresses in geocoding tasks.

The experimental results here reveal that our method outperformed Google Geocoding, thus demonstrating our method's usefulness. In addition, we successfully built a prototype interface that displays relatives on a map.

Future work includes the following. First, we need to improve our method to work effectively in earlier eras. Second, we need to evaluate other kinds of places like permanent addresses. Third, the map interface still has much room for improvement, such as adding the ability to display the relationships of relatives.

6 Conclusions

We proposed a method of converting old addresses to current addresses in order to use geocoding to display people on a map. The approach of our method follows three stages: (1) collecting postal codes relevant to an old address from the Web, (2) judging the best postal code by considering the frequency of each obtained postal code and the edit distance between the old address and an address associated with a postal code, and (3) refining the address by using location data provided by the government. The experimental results show that our method is superior to Google Geocoding API and thus useful for geocoding tasks. Furthermore, a prototype system was implemented as an interface to display people on a map.

References

1. http://www.post.japanpost.jp/zipcode/download.html
2. http://nlftp.mlit.go.jp/isj/
3. https://developers.google.com/maps/documentation/geocoding/
4. Murakami, H., Zheng, N.: Visualizing Family Trees: development of a dynamic family retrieval system. In: CSCW 2012 5th Workshop on Personal Information Management, PIM 2012 (2012)
5. https://developers.google.com/maps/web/
6. Kodansha International Ltd.: Japanese History: 11 Experts Reflect on the Past Large Print Edition. Kodansha International Ltd., Tokyo (2003)
7. https://familysearch.org/
8. Yamamoto, M., Takahashi, Y., Iwasaki, H., Oyama, S., Ohshima, H., Tanaka, K.: Extraction and geographical navigation of important historical events in the web. In: Kim, K.-S., Fröhlich, P., Tanaka, K. (eds.) W2GIS 2011. LNCS, vol. 6574, pp. 21–35. Springer, Heidelberg (2010)