The address matching method of malfunction service worksheet of power customers based on standard address base

Xin Xu^{1,2}, Jun Fu^{1,2}, Zhijie Sun^{1,2}, Xuemei Li^{1,2}

1Electric Power Research Institute of State Grid JIBEI Electric Power Company, Beijing

2.North China Electric Power Research Institute CO.,LTD., Beijing 100045

Abstract. Processing methods stay at the manual processing level, which is inefficient. By analyzing the address of 95598 worksheet of power customers, a standard address base structure and matching rules for storing standard data sets are established, and a matching method of 95598 incident worksheet of power customers based on standard address base is proposed. Drawing on the segment of standard address base, the method also defines the maximum length of the forward matching algorithm and can match along the self-defined address matching rules. It reduces the number of matches between the address to be matched and the standard data set, decreases the target data set used in the next participle, and improves the matching efficiency. In addition, by defining ambiguous addresses and expanding the rule tree, the matching success rate of the addresses to be matched and the flexibility of system implementation are improved

1 Introduction

At present, the matching for power address information in the 95598 incident worksheet of power customers still stays in the manual analysis level. When the user is dissatisfied with the number of power outages and makes a complaint, the operator can only query for the number of incident blackouts and planned blackouts caused by the responsibility of the power supply enterprises within two months in the area through the system to determine whether there are frequent blackouts. Manually checking the number of blackouts is not only inefficient and poorly regulated, but also demands very high staff experience.

In order to solve the existing problems, this paper presents the address segment of 95598 incident worksheet based on standard address data base. The algorithm performs address matching in the standard address base when making the address segment by using the maximum forward matching algorithm. With the help of each segment, the standard address base is searched to obtain the word length of the maximum positive matching algorithm and the address matching rule tree is referenced in real time to achieve the purpose of continuously updating the matching word length and reducing the target data set. Finally, after matching the standard address, terminate the algorithm, return the target data set, complete the standard address output, and then efficiently and accurately realize address matching.

2 Introduction the address matching of 95598 incident worksheet of power

customers based on standard address base

2.1 The Introduction of Matching Method and Framework

The realization of the matching method: first, build a standard database. Standard data base is processed, which is in the administrative area and business area of the State Grid Jibei Electric Power Co., Ltd, to create the address structure data table. Second, address matching. Call the segment algorithm for automatic matching, if the matching is successful, output directly after the conversing format, if the match fails, then output to the library to be processed and wait for manually correction; artificially analyze the reasons, amend and improve the standard database or add ambiguous table data, standardize the address and achieve standardized address output.

2.2 The process to achieve the matching program

Matching program to achieve the process includes three parts.

(1) Establish standard address base.

The standard address mainly provides the standard word length and matching value for the segment matching. Therefore, it is necessary to analyze the address structure of the current fault address and the power outage information, clarify the division of each administrative region, and then construct the corresponding standard data table hierarchically.

* Corresponding author: 676269930@qq.com

After analyzing more than 60,000 malfunction service worksheets and more than 40,000 power outage information in the northern region of Hebei from 2016 to 2017, the current failure address information is the structure composed of province, city, district / county, township / sub-district offices, village / community. The address structure of power outage information is the power unit and power outage range. Among them, that the power supply unit is regarded as part of the address is because some address of the power outage information only provides the districts, counties and villages. The power supply unit is regarded as a reference address information to avoid district identification error of name duplication. The address information within the range of power outage is mainly city, district / county, township / town / sub-district office and village / community. Therefore, the address hierarchy built for the above data structure is shown in Figure 1:

Fig.1. address hierarchy

There into, province to city, city to district / county, district / county to township / town / neighborhood office, township / town / neighborhood office to community, district / county to electricity supply unit are all one-to-many relationship. Through the above address structure analysis, build the following database table, as shown in Figure 2:

Fig.2. structure of standard address base table

(2) The definition of address matching

Taking into account the problem of writing form of 95598 incident worksheet address of power customers, in order to improve the matching efficiency and facilitate matching the format of the current address, this paper sorts out address information in the malfunction service worksheet and summarizes all writing format of the address, which are shown in Table 1:

Table 1 The information of address format

Address classification	The format of address information
Address format of incident worksheet	Province, city, district / county, village / community
	Province, city, district / county, township / town, village / community
	Province, city, township / town, village / community

In order to make representation easy, the tables in the standard address base are numbered (as shown in Table 2), and then the matching rule of the incident worksheet address is defined by numbers (as shown in Table 3). Take the rule 1 in table 3 as an example. When matching the addresses, the data in the provincial table is matched and calculated first. After the provincial table matches successfully, it matches with the city table then and matches them in turn, terminates the operation after the matching is completed, and returns to the standard address. However, if the matching is failed by rule 1 when matching district / county (No. 3), the matching method continues to match directly according to the rule3 and runs until the matching is completed. If you encounter multiple branches in the operation, executes in turn according to the sequence by default.

Table 2 comparison of various table number

Table name	Province list	City list	District list	Electricity supply unit	Township/town/neighborhood office	Village / community list
Number	1	2	3	4	5	6

Table 3 the address matching rule of incident worksheet

Rule	Rule 1	Rule 2	Rule 3
The number of match table	12356	1236	1256

(3) Fuzzy address processing

In the daily work, 95598 customer service directly obtains the address information in the power incident worksheet on the basis of user dictation, so some of the address data obtained is vague and incomplete. After analyzing two categories of addresses, fuzzy address can be divided into two categories of matchable fuzzy address and unmatchable fuzzy address. For the matchable fuzzy address, the success rate of matching can be increased by adding some matching rules. Fuzzy address that can be matched is mainly divided into ambiguous addresses and incomplete title of administrative divisions. For these two addresses, the matching algorithm proposes the following method:

First, construct ambiguous address matching table. Through the establishment of the relationship among ambiguous addresses, incomplete title of administrative divisions and the standard address, data tables is made. When the address is matched with the corresponding administrative division and can not be matched in the

standard table of the standard address base, the data with association in the ambiguous address table can be matched, and the result is obtained according to the matching condition. Second, improving manually the content of the ambiguity table. In the process of matching, the ambiguity table gradually improves the matching rules by manually combing after the failure of automatic matching.

Although ambiguity table, as part of the standard address base, has some redundancy in the entire datasheet setting, and it addresses the issue of matching ambiguous addresses and improves the matching success rate.

For example: compared with the standard address "Baiwang Town, Shuangtashan Town, Shuangluan District, Chengde City, Hebei Province", "Baiwang Town, Shuangtashan, Shuangluan District, Chengde City, Hebei Province" lacks the administrative division title of "town", which belongs to the incomplete appellation of administrative division. Through the analysis of the address, when matching the administrative divisions of "town", the information associated with "town" in ambiguity table is matched, so that the matching is successful, and the matching failure will not be caused by the administrative divisional appellation.

3 The description of matching process

The first step :segment process; address to be matched is loaded into the matching process, the maximum word length and matching target set is limited in accordance with the matching rules, and segment is carried out for address to be matched.

The second step: matching process; the segment's address to be matched is matched to the standard address. If the match is successful, intercept the corresponding length of the administrative division according to the standard address base and circular matching in accordance with the rules; if the match is unsuccessful, query the ambiguous addresses table and match them, output the standard address after the match is unsuccessful, and need to query the matching rule tree to redefine the word length and the standard data set after the match is successful. If the matching rule tree does not have this rule, then this address will jump into the manual processing flow.

The third step: the standard address output; if the process is to be executed automatically, after the matching is completed, the successful specification address is directly output. If it is the matching process of manual handling, the staff are needed to analyze the existing problems, At the same time, amend the standard library, ambiguity table, rule tree according to the problems found, and finally terminate process.

Fig.3. flow chart of address matching process

4 Conclusion

(1) Achieve the effective segment of the fault address

The address matching method is to determine the current matching address range from the address structure data table of the standard address database according to the matching rule tree and the last successful matching administrative region, which realizes the design of the multi-level vocabularies. It can solve the problem of too many matching words caused by a single vocabulary in the matching process and the multi-level vocabulary list can be used to minimize the matching range of the standard vocabulary. The fuzzy address matching design uses the association relationship between the address structure list of the standard address database and the data in the ambiguous address matching table to quickly locate the standard address corresponding to the fuzzy address and effectively resolve the fuzzy address matching. In the process of address matching, rules are used to guide the address matching process, which reduces the number of matches and improves the matching efficiency.

(2) Conducive to early warning work

The realization of address normalized process will help the power sector to achieve worksheet data analysis in the address dimensions on the statistical analysis of worksheets. For example: a regional power outage data analysis, a user preference analysis, a certain number of business statistics and analysis. In addition, the extension of the technology will also contribute to realize plan outage information, complaint worksheet information analysis and address process.

For frequent blackouts complaints management and data analysis, the difficulty lies in the address which is filled in irregularly. The method proposed in this paper solves the problem of irregular address, which is encountered in frequent power outage complaint management and data analysis, and lays the foundation for early warning of complaints and advancement of service gateways.

References

1. Cheng Changxiu, Yu Bin. A matching method of the fuzzy Chinese address segment based on the rules [J]. *Geography and geosciences information science*, 2011 (03): 26-29.
2. Ding Hao. An improved algorithm of Chinese segment based on analytic hierarchy process [J]. *Information Technology*. 2016, 190-193.
3. Xiu Chi. Research and realization on the Chinese segment method adapted to different fields [D]. Beijing University of Technology, 2013.
4. Wang Ruilei, Luan Jing, Pan Xiaohua, Lu Xiufeng. An improved Chinese segment forward maximum matching algorithm. *Computer Applications and Software*, 2011, 28 (3): 195-197.
5. Tan Kankan. Chinese address segment and matching methods based on rules [D]. Shandong University of Science and Technology, 2011.
6. Wang Wentao, Mu Xiaofeng, Wang Lingxia. A Chinese Segment Method Based on Feature Embedded Neural Networks [J]. *Journal of Central South University for Nationalities*. 2017, 36 (1): 102-106.
7. Jin Zaiquan, Zhao Zhao, et al. An improved maximum-caret matching algorithm [J]. *Science Technology and Engineering*, 2007, 7 (9): 4761-4764.
8. Wu Shengyuan. A Chinese segment method [J]. *Computer research and development*, 1996, 33 (4): 306-310.
9. Chen Guilin, Wang Yongcheng, Han Kesong, et al. An improved rigid segment algorithm [J]. 2000, 37 (4): 418-424.
10. Zhang Li, Xu Weiran. Research on Chinese segment [J]. *Software*, 2012, 33 (12): 103-108.