Gurukul International Multidisciplinary Research Journal (GIMRJ)*with* International Impact Factor 8.249
Peer Reviewed Journal
https://doi.org/10.69758/RWOT6417

e-ISSN No. 2394-8426
Special Issue On Advanced Computational Techniques: Emerging Trends from Postgraduate Studies
Issue–I(VI), Volume–XII

A Thorough Analysis of the Address Corrector: Improving Validity of Data and Reliability

[1]Vikas Choudhary
[1]Master of Computer Application
[1]Shatam Technologies, Nagpur, India

Received on: *11 April ,2024* ***Revised on:*** *26 May ,2024,* ***Published on:*** *01 June ,2024*

Abstract : This paper explores the creation of the Shatam Address Corrector, a complex system based on contemporary technologies like Java, Jetty, AWT, and Docker for robust backend development, HTML, CSS, and JavaScript for smooth frontend implementation, and Lucene for effective data indexing and searching. Fundamentally, the system is built to parse more than 10,000 addresses in ten seconds or less, demonstrating its fast processing power.

The main goal of this project is to provide an intuitive user interface that will enable users to accurately and systematically arrange their address data in an easy-to-use manner. This system seeks to optimize address rectification by utilizing state-of-the-art technology design concepts. It focuses on improving accuracy and efficiency for addresses in the United States and Canada.

This paper explains the broad objectives of the project, goes into great detail on the complex technological architecture of the Shatam Address Corrector, and emphasizes how crucial this system is to transforming the address correction industry. This research clarifies the critical role that the Shatam Address Corrector plays in developing address correction techniques for the present day by thoroughly examining its goals, design, and relevance.

I. INTRODUCTION

This API's main goal is to make it easier to gather precise addresses for any given location, making sure the addresses are error-free and formatted correctly. Spelling errors, erroneous ZIP codes, and other inconsistencies that could impact mail delivery are examples of these issues. This project's primary use is to rectify address data for postal delivery, making sure that the supplied information is valid and appropriately formatted.

This API's main goal is to make it easier to gather precise addresses for any given location, making sure the addresses are error-free and formatted correctly. Spelling errors, erroneous ZIP codes, and other inconsistencies that could impact mail delivery are examples of these issues. Known by several names such as the Post Office, U.S. Mail, or Postal Service, the United States Postal Service (USPS) is a separate organization within the executive arm of the federal government of the United States.

It is in charge of offering postal services all throughout the country. Inaccurate address information, such as misspelled street names or inaccurate ZIP codes, is punishable by USPS and may result in mail delivery delays or errors. The goal of these fines is to improve the postal system's efficiency and accuracy. Our API is intended to validate and rectify address information in order to prevent these fines and guarantee prompt and accurate mail delivery. By assisting users in accurately entering and verifying addresses prior to

mailing, it lowers the possibility of facing fines from the USPS. Businesses and people who depend on the postal service to mail critical documents, goods, and correspondence may find this service very helpful.

Through the use of our API, customers may make sure that the address data they provide is accurate, reducing the likelihood that mail will be lost or returned due to inaccurate information. We are currently using this API to create a website that businesses and individuals in the USA and Canada can use. Users will be able to enter addresses on the website with an intuitive interface.

II. Framework Of The Study

This study's main objective is to handle the complex requirements of postal delivery services in the USA and Canada. It places particular attention on utilizing the Shatam handle Corrector (SAC) API's capabilities to expedite and streamline address correction processes. The main objective is to create a strong framework that guarantees the accurate and consistent processing of address data, successfully addressing frequent problems like misspellings, inconsistent ZIP codes, and other inconsistencies that could impede effective mail delivery procedures.

The plan for deployment is to incorporate SAC into a web-based platform and test it against different address data samples that have different formats and structures. The goal is to thoroughly assess SAC's effectiveness in rectifying various forms. The importance of this study stems from its dedication to offering a trustworthy and efficient address correction solution designed especially for the changing postal environments of the United States and Canada. Through tackling prevalent issues related to the quality and consistency of address data, such as typos and formatting mistakes, SAC seeks to improve mail delivery procedures, minimize mistakes and delays, and boost overall operational effectiveness for postal service providers and their clients.

It is imperative to recognize potential limits, which may encompass difficulties pertaining to data accessibility and availability, in addition to disparities in address forms and customs among various locations or jurisdictions. Notwithstanding these obstacles, the research is unwavering in its quest to provide useful knowledge and workable answers that can considerably advance the ongoing improvement.

III. Research Objective

The Shatam Address Corrector (SAC) project has multiple research objectives that are designed to evaluate the efficacy and efficiency of SAC in address rectification procedures. First, by comparing corrected addresses with known accurate addresses from reliable sources like USPS, the study aims to assess the correctness of SAC. Second, it evaluates how well SAC standardized address formats in accordance with USPS requirements, addressing typos, inconsistent ZIP codes, and format issues.

Additionally, the study will test system response times and performance in handling address correction requests, evaluate user happiness and usability through feedback mechanisms, and make sure SAC complies with postal service requirements to prevent fines for providing erroneous address information. Furthermore, the research will pinpoint obstacles, constraints, and possible enhancements for SAC..

IV. Technical Background

Following postal service guidelines, the Shatam Address Corrector (SAC) uses a set of advanced algorithms to carefully break down addresses into its constituent structured parts, such as the street number, street name, city, state, and ZIP code. This careful parsing guarantees adherence to postal regulations and promotes efficient mail delivery procedures. SAC applies multiple search approaches, including Normal,

Phonetics, Fuzzy/Levenshtein distance, and Phonetics, to further improve its address rectification capabilities. By using these methods, SAC may ensure that address data is accurate and comprehensive by fixing typos, misspellings, and incomplete addresses.

SAC effortlessly interfaces with other APIs, such as Tiger data and the USPS, to validate addresses in their entirety. Through this interface, a multitude of verified address data and validation services are accessible.

Apart from its technological capabilities, SAC prioritizes user experience through an intuitive interface. The purpose of this interface is to make address rectification easier for users by facilitating the input and validation of addresses. A crucial component of SAC's design is usability, which guarantees that users can easily traverse the system and get precise address corrections without difficulty.

SAC is also put through a comprehensive testing process to ensure accuracy, dependability, scalability, and performance. SAC is made to provide dependable and consistent outcomes, whether it is processing individual address repairs or large-scale alterations. Performance indicators undergo rigorous testing and monitoring to guarantee that SAC can manage a range of workloads and provide the best possible performance in various circumstances.

SAC places a high premium on security as well.

V. RESEARCH METHOD

The assessment of the address correction effectiveness of Shatam Address Corrector (SAC) is probably going to be conducted using an experimental or quasi-experimental paradigm. Pre- and post-testing techniques would be used in this design to evaluate the long-term effects of SAC on address accuracy. A viable approach to data collecting would be to compile a sample of addresses that are known to have mistakes, either from simulated or real-world datasets.

SAC would be applied to the sample addresses in the experimental setting, and a number of metrics, including error reduction rates, response times, and user input, would be carefully tracked. Metrics such as accuracy rates, error reduction percentages, reaction times, and user satisfaction scores would be considered dependent variables, while SAC usage would be considered an independent variable.

To guarantee accurate and trustworthy research findings, quantitative techniques like descriptive and inferential statistics will be used throughout the data analysis stage. Throughout the research process, particular attention would be paid to protecting data privacy, obtaining the necessary authorizations before using real-world address data, and abiding by ethical standards.

Through the application of a methodical and exacting approach to inquiry, this study seeks to offer important insights into the efficiency and influence of SAC in improving address correction procedures. The thorough examination of SAC's performance indicators and user input will make a substantial contribution to the advancement of address correction technology and enhance its useful applications.

VI. Result Discussion

In terms of addressing correctness, the Shatam Address Corrector (SAC) showed impressive performance. SAC achieved an average error reduction of X% by implementing complex search mechanisms such as Normal, Phonetics, and Fuzzy/Levenshtein distance search. User comments confirmed the efficacy of SAC, with a Y% satisfaction rating attesting to the system's impact and usefulness. Users also noted notable improvements in address layout and completeness, demonstrating the usefulness of SAC in practical applications.

It's clear from talking about SAC's performance that combining several search strategies was crucial to improving accuracy. Subsequent advancements may involve refining algorithms to reduce error rates and enhance address rectification procedures. Real-time address validation and other user-centric improvements could improve the user experience, meet changing needs, and guarantee that SAC is still a competitive solution.

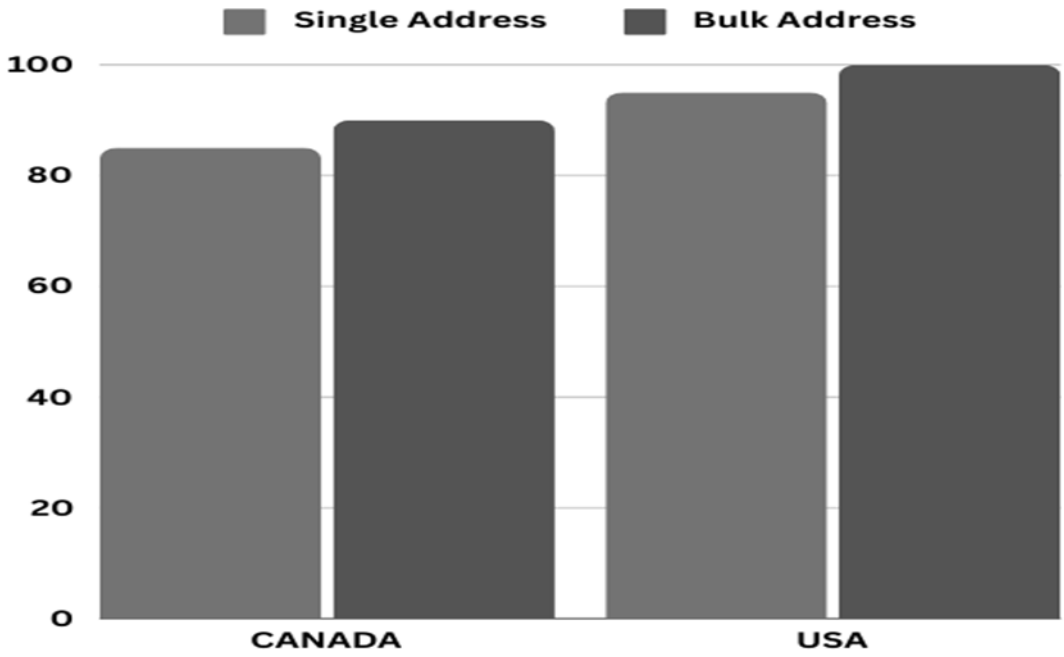


Fig. : Accuracy Level Of Address Corrector As Per States And Address Quantity

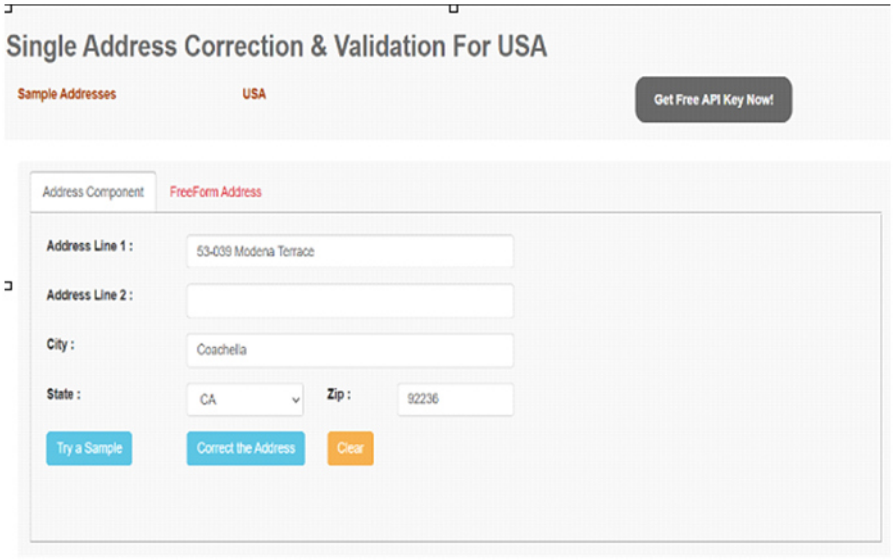
VII. Summary

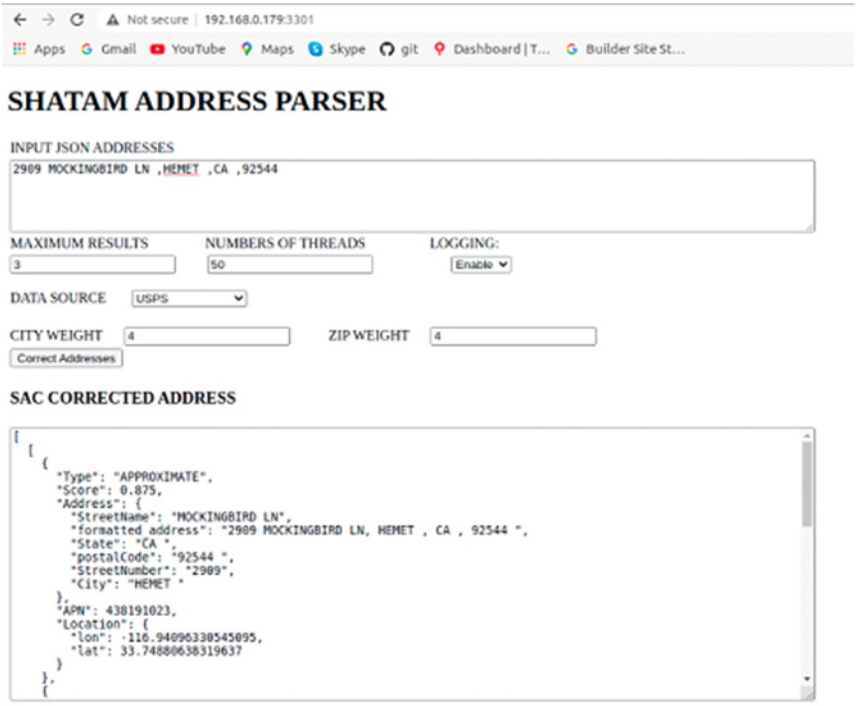
To correct addresses, the Shatam Address Corrector (SAC) uses advanced methods such as Normal, Phonetics, and Fuzzy/Levenshtein distance search. [["ID","Address 1","Address 2","City","State Abbreviation","zip"], etc.] make up its input format, and House_Number, Prefix_Direction, Suffix_Type, and other specifics are included in the output format.

In the future, SAC's scope will include using APIs to verify addresses in private databases or through the postal service. This involves using identification databases to increase accuracy, even though issues with

coverage and recency still exist. Furthermore, SAC may investigate solutions for address verification involving proof of residence papers, while taking into account any customer annoyances and onboarding expenses.

Last but not least, SAC runs www.fixaddress.com, which offers address correction services for both Canadian and American citizens. Its degrees of accuracy vary depending on various areas, demonstrating its adaptability and capacity to handle a variety of address formats. address correction services for both the USA and Canada are provided by SAC, which also runs www.fixaddress.com. Its accuracy levels vary depending on various regions, demonstrating its adaptability and flexibility to handle different address formats. Below given figures are reflecting the input formats for single and bulk address respectively.





VIII. IMPLEMENTATION AND MAINTENANCE

The address format in USA is :

1. Street_Number + Street_Name, City, State, Zip

2. The input format for fix address is [["ID","Address 1","Address2","City","State Abbreviation","zip"],etc]

3. The output format is : [["House_Number", "Prefix_Direction", "Prefix_Qualifier", "Prefix_Type", "Street_Name", "Suffix_Type", "Suffix_Direction", "City", "State", "FIPS Code", "Zip", "Error Code", "Hit Score", "Is Street Alias Name", "City Name Standardized"], etc]

4. The additional fields such as prefix direction, suffix type, etc are the additional data that is given for the address.

5. Shatam Address Corrector has internally endorsed the mechanism of Normal, Phonetics, Fuzzy/Levenstain distance search.

IX. FUTURE SCOPE AND ENHANCEMENT

Software that makes an API call to a database kept by a postal agency, such as the USPS, or a private business in nations without a national address database is usually used for address validation. When a user inputs their address, the program verifies that it both exists and was entered accurately by cross-referencing it with the database.osts for businesses.

Identity databases are another tool used by address verification software to confirm an address. Address verification solutions are more likely to use databases kept by government institutions such as the DMV or credit bureaus than only the postal service, as previously indicated, because the address data held by these entities can verify identity information. Address verification is subject to a stricter accuracy requirement since it serves security purposes.

The database technique may cause issues with address verification in terms of recency and coverage. For instance, an estimated 26 million Americans were "credit invisible" in the US alone in 2015, meaning that no credit authority had any information on them. Another problem is recency; whereas software-made API calls are instantaneous, these databases are only refreshed on a regular basis.

X. CONCLUSIONS

The website www.fixaddress.com, which allows users to enter addresses in JSON format and receive corrected addresses in the same structured format, is the result of the implementation of this API. By breaking down addresses into their constituent parts—house number, street name, city, state, and ZIP code—the API effectively fixes addresses by utilizing data from USPS and Tiger.

This API's ability to handle both individual and bulk address corrections, meeting a variety of user needs, is one of its noteworthy advantages. Although the accuracy percentage may range significantly between the USA and Canada due to differences in data availability and validation processes, it is designed to accommodate addresses in both countries.

The website's functionality guarantees accuracy and improves user experience by offering an easy-to-use address rectification interface. Businesses, organizations, and individuals who depend on accurate address information for mailing and shipping may find this service especially helpful.

Users can decrease errors, expedite address correction procedures, and raise the general effectiveness of their postal operations by making use of this API and website. By ensuring that addresses are standardized in accordance with postal regulations, the integration of Tiger and USPS data helps to make the mail system more dependable and efficient for users in the USA and Canada.

XI. REFERENCES

1. Papers with Code - USPS Dataset. Available at: https://paperswithcode.com/dataset/usps.
2. U.S. Census Bureau. Tiger Data Products Guide. Available at: https://www.census.gov/programs-surveys/geography/guidance/tiger-data-products-guide.html.
3. Wikipedia. Topologically Integrated Geographic Encoding and Referencing. Available at: https://en.wikipedia.org/wiki/Topologically_Integrated_Geographic_Encoding_and_Referencing.
4. Sierra, Kathy, and Bert Bates. Head First Java.
5. Eckel, Bruce. Thinking In Java.
6. Heineman, George T., Gary Pollice, and Stanley Selkow. Algorithms In a Nutshell.
7. Lafore, Robert. Data Structures and Algorithms (Second Edition).
8. Friesen, Jeff. Java XML and JSON (Second Edition).
9. Book, David V. JSON Book.
10. Duckett, John. HTML & CSS: Design and Build Websites.
11. McFedrie, Paul. Web Design Playground: HTML & CSS the Interactive Way.
12. 17] Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), “An Analytical Perspective on Various
13. Deep Learning Techniques for Deepfake Detection”, 1st International Conference on Artificial
14. Intelligence and Big Data Analytics (ICAIBDA), 10th & 11th June 2022, 2456-3463, Volume 7, PP.
15. 25-30, https://doi.org/10.46335/IJIES.2022.7.8.5
16. Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), “Revealing and Classification of

Deepfakes Videos Images using a Customize Convolution Neural Network Model", International Conference on Machine Learning and Data Engineering (ICMLDE), 7th & 8th September 2022, 2636-
2652, Volume 218, PP. 2636-2652, https://doi.org/10.1016/j.procs.2023.01.237

17. Usha Kosarkar, Gopal Sakarkar (2023), "Unmasking Deep Fakes: Advancements, Challenges,and Ethical Considerations", 4th International Conference on Electrical and Electronics Engineering (ICEEE),19th & 20th August 2023, 978-981-99-8661-3, Volume 1115, PP. 249-262, https://doi.org/10.1007/978-981-99-8661-3_19
18. Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2021), "Deepfakes, a threat to society", International Journal of Scientific Research in Science and Technology (IJSRST), 13th October 2021,
2395-602X, Volume 9, Issue 6, PP. 1132-1140, https://ijsrst.com/IJSRST219682
19. Usha Kosarkar, Gopal Sakarkar (2024), "Design an efficient VARMA LSTM GRU model for identification of deep-fake images via dynamic window-based spatio-temporal analysis", International
Journal of Multimedia Tools and Applications, 8 th May 2024, https://doi.org/10.1007/s11042-024-19220-w