iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4669

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Optimized Search Method Using Lazy Search Algorithm

S. Amirtharaj[1], G Saranraj[2], E Balraj[3]
[1, 2]*UG Student,* [3]*Assistant professor, Department of Information Technology, M.Kumarasamy college of Engineering, Karur, India.*

Abstract: ***A Lazy Search is process of the search method will produced optimized solution using Artificial Intelligence. For implementing Lazy Search Algorithm the five technologies are used. The following are technologies which is used for implementing the Lazy Search Algorithm. 1) Mechanize, 2) Beautiful Soup 4, 3) Web Scraping, 4) Python 5) Django. This project seeks to analyze the impact of SEO techniques on the effectiveness of SEO, to figure out which technique strategy is most effective, and furthermore, to test the possible influence of SEO techniques on Page Interest. In this proposed method, the search method is not based on the page ranking algorithm. Lazy search algorithm performance is fully based on the keyword search. Lazy search algorithm will produce the result based on the search keyword and using the key word the search method will produce the optimized solution. Search efforts are constrained by the confidentiality of the algorithm used by the search method themselves.***

Keywords: Mechanize, Beautiful Soup 4, Artificial Intelligence, Web Scraping.

I. INTRODUCTION

In this Search method using Lazy search algorithm is necessary to produce the optimized solution for given keyword. For example, The person want to fix the bug in program for that the person should use the search engines for get the solution for the bug. But the search engine will produce the solution using page rank algorithm or key word search algorithm. This project presents the optimized solution for the given key word using Artificial Intelligence (AI).

II. CONCEPTUAL BACKGROUND

A. Artificial Intelligence

Artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals. The field was founded on the claim that human intelligence "can be so precisely described that a machine can be made to simulate it". This raises The traditional problems (or goals) of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception and the ability to move and manipulate objects. In the twenty-first century, AI techniques have experienced a resurgence following concurrent advances in computer power, large amounts of data, and theoretical understanding; and AI techniques have become an essential part of the technology industry, helping to solve many challenging problems in computer science, software engineering and operations research.

B. Mechanize

Mechanize is programmatic web browsing, after Andy Lester's Perl module WWW:: Mechanize. Mechanize Browser implements the urllib2.OpenerDirector interface. Browser objects have state, including navigation history, HTML form state, cookies, etc. The set of features and URL schemes handled by Browser objects is configurable. The library also provides an API that is mostly compatible with urllib2: your urllib2 program will likely still work if you replace "urllib2" with "mechanize" everywhere. Features include: ftp:, http: and file: URL schemes, browser history, hyperlink and HTML form support, HTTP cookies, HTTP-EQUIV and Refresh, Referrer [sic] header, robots.txt, redirections, proxies, and Basic and Digest HTTP authentication. Much of the code originally derived from Perl code by Gisle Aas (libwww-perl), Johnny Lee (MSIE Cookie support) and last but not least Andy Lester (WWW::Mechanize). urllib2 was written by Jeremy Hylton.

C. Beautiful Soup 4

Beautiful Soup is a library that makes it easy to scrape information from web pages. It sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree. Send a HTTP request to the URL of the webpage you want to access. The server responds to the request by returning the HTML content of the webpage. For this task, we will use a third-party

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

HTTP library for python requests. Once we have accessed the HTML content, we are left with the task of parsing the data. Since most of the HTML data is nested, we cannot extract data simply through string processing. One needs a parser which can create a nested/tree structure of the HTML data. There are many HTML parser libraries available but the most advanced one is html5lib. Now, all we need to do is navigating and searching the parse tree that we created, i.e. tree traversal. For this task, we will be using another third-party python library, Beautiful Soup. It is a Python library for pulling data out of HTML and XML files.

D. Web Scraping

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. Web scraping software may access the World Wide Web directly using the Hypertext Transfer Protocol, or through a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a bot or web crawler. It is a form of copying, in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis.Web scraping a web page involves fetching it and extracting from it. Fetching is the downloading of a page (which a browser does when you view the page). Therefore, web crawling is a main component of web scraping, to fetch pages for later processing. Once fetched, then extraction can take place. The content of a page may be parsed, searched, reformatted, its data copied into a spreadsheet, and so on. Web scrapers typically take something out of a page, to make use of it for another purpose somewhere else. An example would be to find and copy names and phone numbers, or companies and their URLs, to a list (contact scraping). Web scraping is used for contact scraping, and as a component of applications used for web indexing, web mining and data mining, online price change monitoring and price comparison, product review scraping (to watch the competition), gathering real estate listings, weather data monitoring, website change detection, research, tracking online presence and reputation, web mash up and, web data integration.

E. Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. Van Rossum led the language community until July 2018. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python features a comprehensive standard library, and is referred to as "batteries included".Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open-source software[30] and has a community-based development model. Python and CPython are managed by the non-profit Python Software Foundation.

F. DJANGO

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It’s free and open source. Django includes dozens of extras you can use to handle common Web development tasks. Django takes care of user authentication, content administration, site maps, RSS feeds, and many more tasks — right out of the box. Django takes security seriously and helps developers avoid many common security mistakes, such as SQL injection, cross-site scripting, cross-site request forgery and clickjacking.

III. IMPLEMENTATION

The search algorithm will work based on the key word. The user will provide the key to the text box. The given text will redirect to the mechanize tool. The tool will fetch the website based on the keyword. Then the BS4 fetch the html content and it will modify the html content. BS4 is use to beautify the html content. From the html content the web scrapping will perform to fetch the particular content of the result. Artificial intelligence provide the exact answer for the given keyword. Then display the result to the user.

©IJRASET: All Rights are Reserved

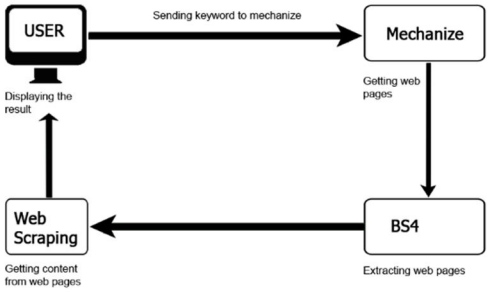


Fig 1.1 Architecture of Lazy search algorithm

The above architecture diagram will describe the process of the lazy search algorithm.

A. Algorithm

1) Input: keyword
2) Output :optimized result for the keyword
a) Begin
b) keyword ← user input;
c) mechanize ← keyword;
d) fetch the website;
e) Beautiful Soup 4 ← website link;
f) Extract the content;
g) Beautify the content;
h) web scrapping ← BS4 content;
i) scrap the content;
j) write the html content from scrapping text;
k) resutl ← html content;
l) end

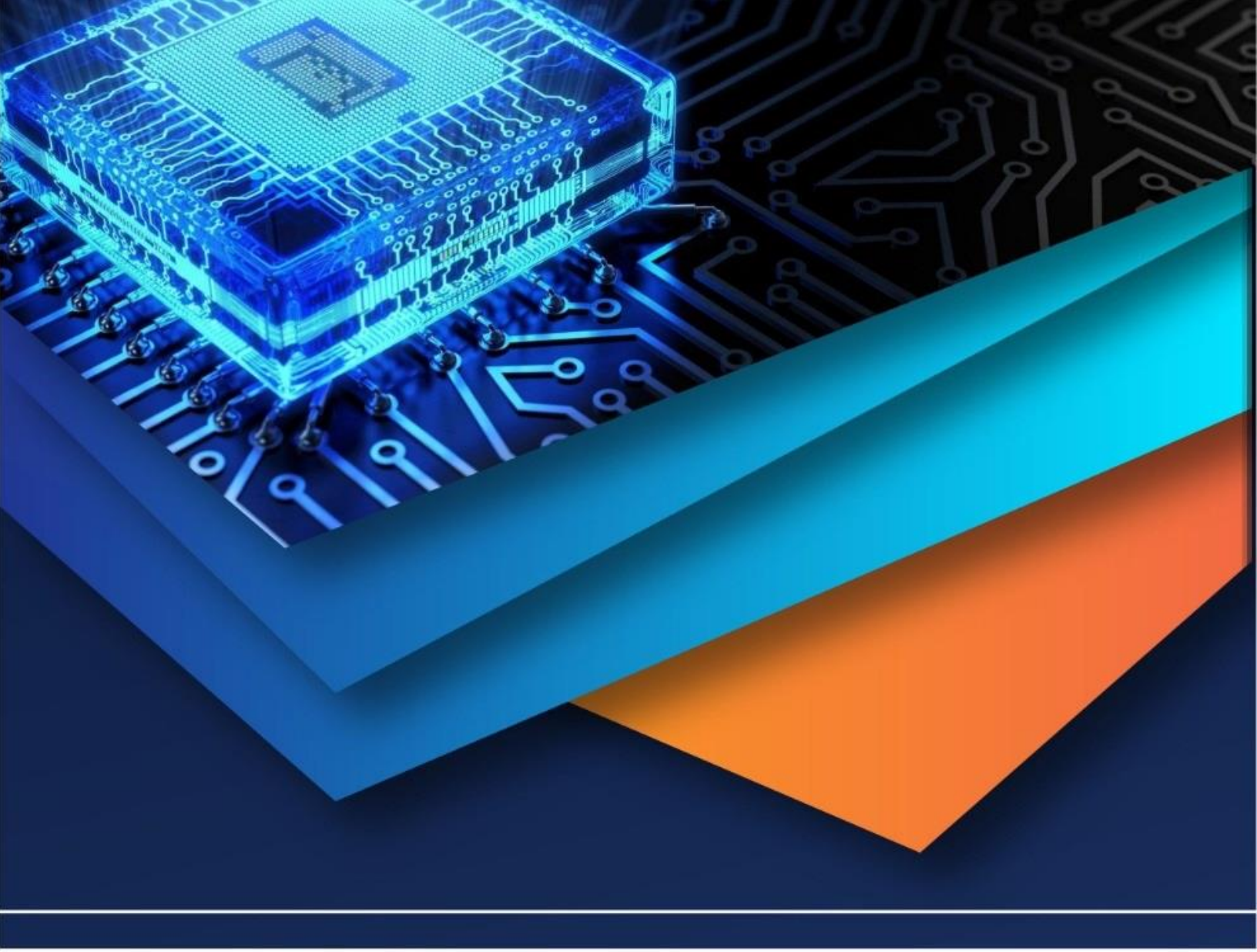
IV. CONCLUSION

While using optimized search method using lazy search algorithm the accurate result will be generated for the given keyword. Due to this the search time will be reduced, search operation will be efficient. Using this lazy search algorithm the user can get the result in optimized method. For that the user will get the good experience and result.

REFERENCE

[1] J.G. Hazan, “Stop Being Evil: A Proposal for Unbiased Google Search,” Michigan Law Review,vol. 111, no. 5, pp. 789–820, 2013.
[2] S.K. Ganta and S.P.K Somayajula, “Search Engine Optimization through Spanning Forest Generation Algorithm,” Int. J. on Computer Science and Engineering, vol. 3, no. 9, pp. 3275–3282,Sep. 2011.
[3] S. Rawat, R. Chowdhary, and D.A. Bansal, “Data Integrity of Cloud Data Storages (CDSs) in Cloud,” Int. J. Advanced Research in Computer Science and Software Engineering, vol. 3, no. 3, pp. 588– 593,Mar. 2013.
[4] J.M. Casarella, “The Application of Hierarchical Temporal Memory to the Evaluation of EEG Signals,” Proc. of Student/Faculty Research Day 2007, School of Computer Science and Information Systems, Pace University, New York (NY, USA), May 2007.
[5] J.D. Neushul, “Interoperability, Data Control and Battlespace Visualization Using XML, XSLT and X3D,” Master Thesis, Department of Computer Science, Naval Postgraduate School, Monterey (CA, USA),Sep. 2003.
[6] V. Nourani, A.H. Baghanam, J. Adamowski, and O. Kisi, “Applications of Hybrid Wavelet–Artificial Intelligence Models in Hydrology: A Review,” J. Hydrology, vol. 514, pp. 358–377, Jun 2014,doi: 10.1016/j.jhydrol.2014.03.057.
[7] G. Yu and Z. Schwartz, “Forecasting Short Time-Series Tourism Demand with Artificial Intelligence Models,” J. Travel Research, vol. 45, no. 2, pp. 194–203,Nov 2006, doi: 10.1177/0047287506291594.
[8] A.A. Antonov, “From Artificial Intelligence to Human SuperIntelligence,” Int. J. Computer Information Systems, vol. 2, no. 6, pp. 1–6, Jun. 2011.
[9] A. Kumar, “Search Engine Optimization (SEO): Technical Analysis Concepts,” Int. J. Emerging Technology and Advanced Engineering, vol. 3, no. 3, pp. 123–128,Mar. 2013.
[10] J.G. Hazan, “Stop Being Evil: A Proposal for Unbiased Google Search,” Michigan Law Review,vol. 111, no. 5, pp. 789–820, 2013.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)