BUZZMATCH: FILE CONTENT-BASED WITH IMAGE AND TEXT FILTERING CATEGORIZATION APPLICATION THROUGH KEYWORD MATCHING USING MODIFIED ARTIFICIAL BEE COLONY (ABC) APPROACH

A Thesis

Presented to the

Faculty of the

College of Science

Technological University of the Philippines

Manila

In Partial Fulfillment of the

Requirements for the Degree

Bachelor of Science in Computer Science

By:

ESPINOSA JOESEF ANDREI

RIGA, RASHEED

TORRES, JAN CHRISTIAN

VALLOYAS, JON REXZEL

June 2024

INTRODUCTION

The File Content-Based with Image and Text Filtering Categorization Application Via Keyword Matching using Modified Artificial Bee Colony (ABC) Approach. The ABC algorithm is applied utilizing a more efficient way to automate the categorization of the electronic copious number of text-based documents. The study has the following specific objectives: Determine the acceptability level of the system using the ISO 25010 criteria in functional, performance efficiency, usability, reliability, and maintainability. The system is able to dynamically adjust itself according to the features and difficulties of e-texts. The use of electronic text files in organizations and industries has increased steadily. This poses a problem about how to categorize these files according to their content. Using automated text-based categorization can help professors and teachers save time finding resources on their subjects. The system focuses on utilizing the Artificial Bee Colony (ABC) as the core technique for the categorization. The outcomes of this study are expected to contribute to the body of knowledge, enhancing understanding of these records based on what they contain. The researchers have come up with a solution that uses AI for automation. Paper documents have been the primary tool for storing and sharing information. Electronic text-based documents have features like searchable texts, bookmarks and hyperlinks. Organizations can overcome the challenges involved in manual categorization by developing artificial intelligence (AI) systems. The significance of PDF?s versatility in Python has been discussed in the past. The ABC algorithm has been used to optimize the categorization process, which is inspired by Bee Foraging behavior. This kind of protection is very essential for a number of industries.

METHOD

BuzzMatch is designed to be a valuable tool for students, faculty, and professionals who need to organize and organize documents. 30 respondents will be selected, comprising 15 non-IT professionals, and 15 IT/CS professionals, using convenience sampling. BuzzMatch leverages the concept of parallel processing in bee colonies to efficiently distribute the categorization across multiple CPU cores. The data will be organized in an Excel file to calculate mean scores. The BuzzMatch Likert Scale was employed to interpret the adjectival scores. The development of an application involves several stages. The first step is to select tools and frameworks to meet the requirements. This is followed by the creation of a mockup, which showcases prototype. The application is then uploaded to the server for beta testing or real use. The final step is the creation and deployment of the application. The software is then ready for use by the user. It can be downloaded from the server or downloaded to the user's computer. It is then available for download.

RESULTS

The respondents evaluated the Reusability, Fault Tolerance, Learnability, and Operability categories. They also evaluated the Time Behavior, Resource Utilization, and Performance Efficiency categories. The results were published in a paper published by the European Computer Society (ECS) in September 2013. The paper was based on a survey of more than 1,000 IT professionals. It was published by the European Computer Society, a member of the European Commission (EC) and the European Association of IT Professionals (ECA). The system has the following capabilities: OCR, Press Categorization, and Analyzability. The system is designed to be used by IT/CS professionals. The OCR Module contains information such as select input and output folders. The Text Categorizer Module contains the ability to categorize textbox/folders. There is a dropdown on how many categories or folders the user wants to use. The Analyzable Module allows the user to search for keywords in a particular category. The evaluation for the system received a favorable response for both IT/CSprofessionals and Non-IT/CS professionals. The system is designed to handle only TXT, PDF and DOCX files. The application includes error handling mechanisms to manage incorrect or missing user input. The keyword matching currently currently uses an "AND" logic, meaning a document must contain keywords in a category to be assigned to that category. There are three categories, Functional Suitability, Reliability and Maintainability. The table 9 shows the Overall Summary of Responses in different evaluation. The average score is 3.53 Highly Acceptable. The overall score is 3.6 Highly Accepted. The average score is also 3.6 High Acceptable, the highest score is 4.0 High Acceptability, and the lowest score is 2.5 High Acceptance. For more information on the evaluation process, see the http://www.cnn.com/2013/01/29/technology/top-10-researchers-in-the-world-2013/index.html.

DISCUSSION

Determine the acceptability level of the developed system using the ISO 25010 criteria. Test and evaluate the system in terms of functional suitability and reliability. Integrate the text categorizer system with the Optical Character Recognition (OCR) for enhanced document analysis and categorization. Use the evaluation to identify areas for further improvement in the scheduling system. The recommended enhancements for the development of the application are as follows: Python, Tkinter, and the OCR module for TkInter.