ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

Design and Implementation of Domestic News Collection System Based on Python

[1] K SUPARNA, [2] S. YASWANTH RAM

[1](Assistant Professor), MSC, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MSC, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

The rapid development of the Internet, network media has become a new window for people to understand the outside world due to its fast speed and wide spread. News is a channel for people to know about Surrounding Information, but thousands of news are produced every day on the Internet. These news are needed or not in inside. How to efficiently and accurately obtain the news content we need from the website is a great need in people's life.This system aims to collect news on specific websites and return it to users with concise and clear pages. Users can search specific keywords to select news that they are interested in so as to realize personalization for users. This system crawls and processes the domestic financial news content, which is convenient for people to process the information. In order to avoid duplication of information, the system has also implemented a self-defined deduplication rule. In the specific implementation, the system is written using Python in conjunction with the Scrapy framework and Django framework, which can simplify the system code to a certain extent. The practical value of this system lies in the timely, efficient and convenient access to domestic financial news that people care about, need and are interested .

1.INTRODUCTION

News is an important way to convey information. Among the tens of thousands of news generated every day, obtaining effective news is an important objective. How to get news conveniently and efficiently has become an important orientation. Nowadays, a full-featured news-gathering platform has become more and more popular and has good development prospects [1].This paper designs and develops a convenient automatic news-gathering system. The system uses crawler analysis to collect domestic news, saves it after deduplication, and finally provides news services for retrieving and viewing. It can help users find similar news and extract hot news that users are interested in, and improve the efficiency of readinnews .

News is an important way to convey information. Among the tens of thousands of news generated every day,obtaining effective news is an important objective. How to get news conveniently and efficiently has become an important orientation. Nowadays, a full-featured news-gathering platform has become more and more popular and has good development prospects [1].This paper designs and develops a convenient automatic news-gathering system. The system uses crawler analysis to collect domestic news, saves it after deduplication, and finally provides news services for retrieving and viewing. It can help users find similar news and extract hot news that users are interested in, and improve the efficiency of reading.

2.LITERATURE SURVEY

1)**Design and Implementation of Intelligent News Collection and Processing System**

AUTHORS: J. L. Zhang

Purpose: The goal of XPRESS is to establish a breakthrough for the factory of the future with a new flexible production concept based on the generic idea of "specialized intelligent process units" ("Manufactrons") integrated in cross-sectoral learning networks for a customized production. XPRESS meets the challenge to integrate intelligence and flexibility at the "highest" level of the production control system as well as at the "lowest" level of the singular machine. Design/methodology/approach: Architecture of a manufactronic networked factory is presented, making it possible to generate particular manufactrons for the specific tasks, based on the automatic analysis of its required features. Findings: The manufactronic factory concept meets the challenge to integrate intelligence and flexibility at the "highest" level of the production control system as well as at the "lowest" level of the singular machine. The quality assurance system provided a 100% inline quality monitoring, destructive costs reduced 30%-49%, the ramp-up time for the set-up of production lines decreased up to

50% and the changeover time decreased up to 80%. Research limitations/implications: Specific features of the designed manufactronic architecture, namely the transport manufactrons, have been tested as separate mechanisms which can be merged into the final comprehensive at a later stage.Practical implications: This concept is demonstrated in the automotive and aeronautics industries, but can be easily transferred to nearly all production processes. Using the manufactronic approach, industrial players will be able to anticipate and to respond to rapidly changing consumer needs, producing high-quality products in adequate quantities while reducing costs. Originality/value: Assembly units composed of manufactrons can flexibly perform varying types of complex tasks, whereas today this is limited to a few pre-defined tasks. Additionally, radical innovations of the manufactronic networked factory include the knowledge and responsibility segregation and trans-sectoral process learning in specialist knowledge networks.

2)**Big data method and innovation in news communication:from theoretical definition to operational route**

AUTHORS: G. M. Yu

This article discusses methodological aspects of Big Data analyses with regard to their applicability and usefulness in digital media research. Based on a review of a diverse selection of literature on online methodology, consequences of using Big Data at different stages of the research process are examined. We argue that researchers need to consider whether the analysis of huge quantities of data is theoretically justified, given that it may be limited in validity and scope, and that small-scale analysesof communication content or user behavior can provide equally meaningful inferences when using proper sampling, measurement, and analytical procedures.

3)**Web-based news gathering system**

AUTHORS:J. F. Hu, Y. B. Shen

This study found journalists use government sites most often to retrieve information. Problems include difficulty with verification, unreliable information and lack of contact information.

4)**Keyword extraction algorithm based on automatic text classification**

AUTHORS:H. Zhang

Automatic keyword extraction is an important research direction in text mining, natural language processing and information retrieval. Keyword extraction enables us to represent text documents in a condensed way. The compact representation of documents can be helpful in several applications, such as automatic indexing, automatic summarization, automatic classification, clustering and filtering. For instance, text classification is a domain with high dimensional feature space challenge. Hence, extracting the most important/relevant words about the content of the document and using these keywords as the features can be extremely useful. In this regard, this study examines the predictive performance of five statistical keyword extraction methods (most frequent measure based keyword extraction, term frequency-inverse sentence frequency based keyword extraction, co-occurrence statistical information based keyword extraction, eccentricity-based keyword extraction and TextRank algorithm) on classification algorithms and ensemble methods for scientific text document classification (categorization). In the study, a comprehensive study of comparing base learning algorithms (Naïve Bayes, support vector machines, logistic regression and Random Forest) with five widely utilized ensemble methods (AdaBoost, Bagging, Dagging, Random Subspace and Majority Voting) is conducted. To the best of our knowledge, this is the first empirical analysis, which evaluates the effectiveness of statistical keyword extraction methods in conjunction with ensemble learning algorithms. The classification schemes are compared in terms of classification accuracy, F-measure and area under curve values. To validate the empirical analysis, two-way ANOVA test is employed. The experimental analysis indicates that Bagging ensemble of Random Forest with the most-frequent based keyword extraction method yields promising results for text classification. For ACM document collection, the highest average predictive performance (93.80%) is obtained with the utilization of the most frequent based keyword extraction method with Bagging ensemble of Random Forest algorithm. In general, Bagging and Random Subspace ensembles of Random Forest yield promising results. The empirical analysis indicates that the utilization of keyword-based representation of text documents in conjunction with ensemble learning can enhance the predictive performance and scalability of text classification schemes, which is of practical importance in the application fields of text classification.

5)**Hiding Data in Images Using Cryptography and Deep Neural Network**

AUTHORS:SharmaKartik; Aggarwal Ashutosh;

Steganography is an art of obscuring data inside another quotidian file of similar or varying types. Hiding data has always been of significant importance to digital forensics. Previously, steganography has been combined with cryptography and neural networks separately. Whereas, this research combines steganography, cryptography with the neural networks all together to hide an image inside another container image of the larger or same size. Although the cryptographic technique used is quite simple, but is effective when convoluted with deep neural nets. Other steganography techniques involve hiding data efficiently, but in a uniform pattern which makes it less secure. This method targets both the challenges and make data hiding secure and non-uniform.

3.SYSTEM EXCISTING

Among the tens of thousands of news generated every day,News is a channel for people to know about Surrounding Information, but thousands of news are produced every day on the Internet.How to efficiently and accurately obtain the news content we need from the website is a great need in people's life.

DISADAVANTAGES:

- Low Efficiency.
- We use Large amount of Code.
- Deduplication is not allowed

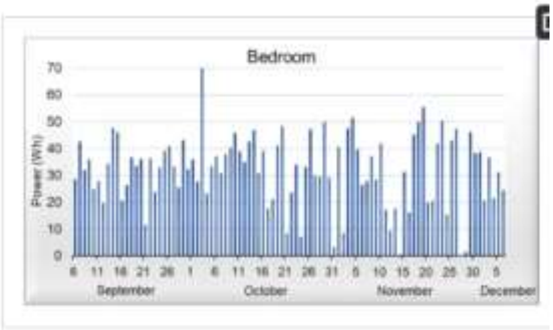
PROPOSED SYSTEM:

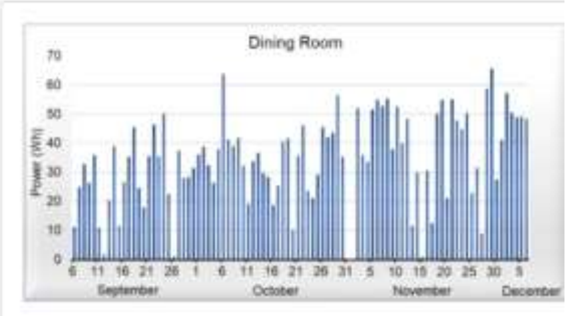
Designs and develops a convenient automatic news-gathering system.The domestic financial news collection system based on python needs to realize the functions of crawling, formatting, storing data, displaying data, operating data (viewing or deleting a news) of various websites.Users can search specific keywords to select news that they are interested in so as to realize personalization for users.Deduplication avoids repeated visits to web pages.

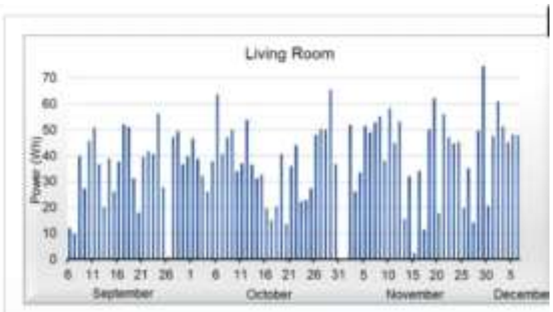
ADAVANTAGES:

- High Efficiency.
- Simplifies the code writing and improves Speed and efficiency of reptiles
- Deduplication is not allowed.

4. OUTPUTSCREENS







5. CONCLUSION

This system makes every effort to facilitate the processing of news information for users, and presents the news information obtained from various websites to the users. The simple and efficient interface enables users to read the news clearly, and only crawls and displays the key information of the news and ignores other unnecessary information, so that users can find the content they are interested in or need more quickly. In short, this system, as a comprehensive information, analysis and retrieval tool, will facilitate people's lives to a certain extent.Certainly, this system can't be perfect, there are still many functions that can be expected, and there are some deficiencies that can be improved. For example, the system currently only implements crawling of a few sites, and the number of crawled sites can continue to be expanded to make news content richer and more complete. Furthermore, if a website is frequently accessed, this website may detect crawlers and block the crawlers. For this problem, you can set a certain anti-crawling strategy to avoid system failure. On the page display, the system can be further optimized to make the interface more concise and intuitive; in the system functions, the functions can be further expanded. These are the goals and directions of this system. This process needs to be optimized step by step to achieve.

6. REFERENCE

HaixiaLv College of Computer Science and Technology Shandong University of Finance and Economics Jinan China xzdjl@126.com,“**Design and Implementation of Domestic News Collection System Based on Python**”, June 16,2020 at 05:45:11 UTC from IEEE Xplore