© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 08 Issue: 05 | April - June 2017
Paper is available at www.jrps.in | Email : info@jrps.in

Review On Bug Detection In Text Based Using Kmp & Bm Algorithm

Ankit Jangra[1], Shilpa Nagpal[2]

[1]Research Scholar, Department of computer science engineering, Prannath Parnami Institute of Management & technology hisar, Ankitjangra92@gmail.com
[2]Department of computer science engineering, Prannath Parnami Institute of Management & technology hisar, Shilpanagpal31@gmail.com

ABSTRACT:- In data mining algorithms could be used, a target data set must be assembled. As data mining could only uncover patterns actually present in data, target data set must be large enough to contain these patterns while remaining concise enough to be mined within an acceptable time limit. The objective of our research is to decrease the time consumption during pattern matching. We have discussion create a function to implement KMP pattern matching using MATLAB and test second step we would create Booyer Moore pattern matching using Matlab and test it.

ISSN : 2278-6848 03
9 772278 684800
© International Journal for Research Publication and Seminar

[1] INTRODUCTION

There is a huge amount of data available in Information Industry. This data is of no use until it is converted into useful information. It is necessary to analyze this huge amount of data and extract useful information from it. Extraction of information is not the only process we need to perform; data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Once all these processes are over, we would be able to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration, etc.

The purpose of data mining is to extract useful information from large databases or data warehouses. Data mining applications are used for commercial & scientific sides. This study mainly discusses Data Mining applications in scientific side. Scientific data mining distinguishes itself in sense that nature of datasets is often very different from traditional market driven data mining applications.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 08 Issue: 05 | April - June 2017
Paper is available at www.jrps.in | Email : info@jrps.in

[2] REVIEW OF LITERATURE

M. O. Mansur, 2 Mohd. Noor Md. Sap 2005 Outlier Detection Technique in Data Mining: A Research Perspective

While field of data mining had been studied extensively, most of work had concentrated on discovery of patterns. Outlier detection as a branch of data mining had many important applications, & deserves more attention from data mining community. Most methods in early work that detects outliers independently had been developed in field of Statistics

Wahidah Husain 2011 Application of Data Mining Techniques for Improving Software Engineering

Computer software plays an important role in business, government, societies, & sciences. To solve real-world problems, it is important to improve quality & reliability of computer software. Software Engineering is computing field concern within designing, developing, implementing, maintaining, & modifying software. Software Engineering data consists of sequences, graphs, & text.. In this paper, we study how data mining techniques could be applied in solving Software Engineering problems..

Ashish Sureka 2011 Detecting Duplicate Bug Report Using Character N-Gram-Based Features we present an approach to identify duplicate bug

Duplicate reports needs to be identified to avoid a situation where duplicate reports get assigned to multiple developers. Also, duplicate reports could contain complementary information which could be useful for bug fixing. Automatic identification of duplicate reports (from thousands of existing reports in a bug repository) could increase productivity of a Triager by reducing amount of time a Triager spends in searching for duplicate bug reports of any incoming report. proposed method uses character N-gram-based model for task of duplicate bug report detection.

V. Neelima1, 2013 Annapurna Bug Detection through Text Data Mining [4]

Text Mining is an interdisciplinary field that draws on information retrieval, Data Mining, machine learning, statistics, & computational linguistics. As most information is currently stored as text, Text Mining is believed to had a high commercial potential value.

[3] PATTERN MATCHING

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 08 Issue: 05 | April - June 2017
Paper is available at www.jrps.in | Email : info@jrps.in

Most people use pattern matching in some form. Search engines on Web use pattern matching to locate information of interest. Patterns could be specific or quite general, using various wildcards that match multiple endings, words, or strings. Many databases have a similar capability, in which a character such as an asterisk is used as a wildcard.

This concept could be extended in many directions. For example, if you wanted to search for information related to "molecules", you might want to include terms such as "molecule" & "molecular". But rather than typing all possibilities, you could add an asterisk & search for "molecul*" to retrieve all three possibilities at once. In Chemical Abstracts Service Molecular Database, researchers could use Markush structures attached to drawings of molecules to retrieve a base molecule, such as benzene, with any string attached, such as OH COOH & Cl. Most bioinformatics databases have similar pattern-matching capabilities. In bioinformatics, flexible pattern matching is called similarity searching. Text mining is an important step of knowledge discovery process. It is used to extract hidden information from not-structured or semi-structured data. They had been developed a crawler for getting pattern matched within a given text by using several algorithms such as:

1. Knuth-Morris-Pratt algorithm(KMP)
2. Boyer-Moore algorithm

After a thorough study a conclusion was drawn that Knuth-Morris-Pratt was paramount choice for our work. & we would make comparative analysis between Finite automata algorithm, Knuth-Morris-Pratt algorithm(KMP), Boyer-Moore algorithm.

[4] PROPOSED WORK

In this research we have to enhance performance of existing pattern matching algorithm to check anomaly by modifying them. objective of our research is to decrease time consumption during pattern matching.

Our proposed work consists of following steps:

Establishment & configuration of cloud environment.

1. We would create a function to implement KMP pattern matching using MATLAB & test it.

2. In second step we would create Booyer Moore pattern matching using Matlab & test it.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 08 Issue: 05 | April - June 2017
Paper is available at www.jrps.in | Email : info@jrps.in

3. study of limitation of KMP & Booyer Moore pattern would be done.

4. Then we would develop a Graphic user interface environment to implement existing KMP & Booyer Moore pattern matching function & get time consumption to perform pattern matching using tic toc function in MATLAB.

5. Then we would develop a new function to implement proposed pattern matching in lesser time.

6. performance chart of existing & new algorithm would be developed.

7. Analysis of performance of proposed pattern matcher would be done using real data set.

8. This proposed algorithm would be integrated to cloud environment to find anomalies.

[5] CONCLUSION

In this research checking of a given sequence of tokens for presence of constituents of some pattern is made to find anomaly in cloud environment. Proposed algorithm would be integrated to cloud environment to find anomalies. Contrast to pattern recognition; match usually has to be exact. patterns generally have form of either sequences or tree structures. Uses of pattern matching include outputting locations (if any) of a pattern within a token sequence, to output some component of matched pattern, & to substitute matching pattern with some other token sequence. Patterns are often described using regular expressions & matched using techniques such as backtracking. BOYER-MOORE ALGORITHM algorithm scans characters of pattern from right to left beginning with rightmost one.

REFERENCE

1. Mohammadjafar Esmaeili (2011) "**Stream Data Mining & Anomaly Detection"** *International Journal of Computer Applications (0975 – 8887) Volume 34– No.9, November 2011* 38
2. Sushil Kumar (2012) "Anomaly Detection in Network using Data mining Techniques" **International Journal of Emerging Technology & Advanced Engineering ISSN 2250-2459, Volume 2, Issue 5, May 2012**
3. **Harshna (2013) "**Survey paper on Data Mining techniques of Intrusion Detection" ***International Journal of Science, Engineering & Technology***

Research (IJSETR) Volume 2, Issue 4, April 2013

4. **Murad A. Rassam (2013)** "**Advancements of Data Anomaly Detection Research in Wireless Sensor Networks" A Survey & Open Issues**
5. **Mohsen Kakavand (2014)** "**A Survey of Anomaly Detection Using Data Mining Methods for Hypertext Transfer Protocol Web Services"**
6. Prashansa Chouhan (2015) "**A Survey: Analysis of Current Approaches in Anomaly Detection"** *International Journal of Computer Applications (0975 – 8887) Volume 111 – No 17, February 2015*
7. Crosbie M. & Spafford G., **"Applying genetic programming to intrusion detection,"** presented at AAAI Fall Symp. Series, AAAI Press, Menlo Park, CA, Tech. Rep. FS-95-01, 1995.
8. Dickerson, J. E. & J. A. Dickerson, **"Fuzzy network profiling for intrusion detection"**, In Proc. of NAFIPS 19th International Conference of North American Fuzzy Information Processing Society, Atlanta, pp. 301306. North American Fuzzy Information Processing Society (NAFIPS), July 2000.
9. Bankovic Z., Stepanovic D., Bojanic S., **"Improving Network Security using Genetic Algorithm Approach"**, Computer & Electrical Engineering, pp. 438-451, 2007.
10. Ektefa M., Memar S., **"Intrusion Detection Using Data Mining Techniques"**,IEEE Trans., 2010.