

THE TITLE

AUTHOR NAME

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS OF THE
UNIVERSITY NAME
FOR THE DEGREE OF
MASTER OF SCIENCE

RESEARCH UNDERTAKEN IN THE
DEPARTMENT OF DEPARTMENT NAME
IN COLLABORATION WITH
COMPANY NAME IN COUNTRY
SUPERVISOR: PROF. APPLE
SECOND SUPERVISOR: PROF. BANANA

MAY 2017

Statement of Authorship

I certify that the attached thesis is my original work. No other person's work has been used without due acknowledgment. Except where I have clearly stated that I have used some of this material elsewhere, it has not been presented by me for examination in any other course or subject at this or any other institution.

I understand that the work submitted may be reproduced and/or communicated for the purpose of detecting plagiarism. I understand that should this declaration be false, I am liable to be penalized under the University Name regulations.

Student's signature _____ Date _____

Acknowledgments

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Author
Author Name

Contents

Statement of Authorship	ii
Acknowledgments	iii
List of Tables	v
List of Figures	vi
1 Introduction	1
1.1 Image	1
1.2 Auto-wrap verbatim text	1
1.3 Source code	2
1.4 Algorithm	2
1.5 Short table	3
1.6 Long table	3
2 State of the art	5
3 Technical solution	6
4 Evaluation	7
5 Conclusion	8
Bibliography	9

List of Tables

1.1	The three log repositories for testing	3
1.2	The full list of use cases	3

List of Figures

1.1	Princeton University	1
1.2	The general bootstrapping algorithm	2

Chapter 1

Introduction

1.1 Image



Figure 1.1: Princeton University

1.2 Auto-wrap verbatim text

```
May 18 04:59:58 server26 sshd[17560]: Accepted publickey  
for fuser from 192.168.174.10 port 39039
```

```
May 18 05:00:02 server24 sshd[20566]: Accepted publickey  
for fuser from 192.168.174.10 port 34449
```

1.3 Source code

```
1 public class HelloWorld {  
2     public static final void main(String[] args) {  
3         System.out.println("Hello, World");  
4     }  
5 }
```

1.4 Algorithm

```
1: seedWords  $\leftarrow$  input ▷ mandatory  
2: seedPatterns  $\leftarrow$  input ▷ optional  
3: newWords  $\leftarrow$  seedWords  
4: allPatterns  $\leftarrow$  seedPatterns  
5: repeat  
6:     allPatterns  $\leftarrow$  allPatterns  $\cup$  generatePatternsFrom(newWords)  
7:     bestPatterns  $\leftarrow$  selectBestSignificantPatterns(allPatterns)  
8:     removeBadPatterns(bestPatterns)  
9:     if sizeof(bestPatterns) = 0 then  
10:         break  
11:     end if  
12:     newWords  $\leftarrow$  extractWordsFrom(bestPatterns)  $\setminus$  seedWords  
13:     bestWords  $\leftarrow$  selectBestWords(newWords)  
14:     seedWords  $\leftarrow$  seedWords  $\cup$  bestWords  
15:     newWords  $\leftarrow$  bestWords  
16: until detectStoppingCondition() = true  
17: newPatterns  $\leftarrow$  allPatterns  $\setminus$  seedPatterns  
18: prune(newPatterns)  
19: seedPatterns  $\leftarrow$  seedPatterns  $\cup$  newPatterns
```

Figure 1.2: The general bootstrapping algorithm

1.5 Short table

Table 1.1: The three log repositories for testing

Name	Content	Size in MB	Number of events
threat-log.csv	A CSV file containing threat logs in 54 columns.	75	200,000
nagios.txt	A non-csv file containing logs in nagios format concerning notifications, alerts and performance data provided by SAP.	100	400,000
syslog.txt	A non-csv file containing standard syslog provided by SAP.	208	2,000,000

1.6 Long table

Table 1.2: The full list of use cases

No.	Description	Input
1	Print the usage of this prototype	
2	List all repositories and tasks with bare info	
3	Create a new repository	repository name
4	Delete a repository and all of its tasks	repository ID
5	Rename a repository	repository ID, new name
6	View content (log events) of a repository	repository ID
7	Add log events (specified in a text file) to a repository	repository ID, file name
8	Remove log events (specified in a text file) from a repository	repository ID, file name
9	Clear all log events in a repository	repository ID
10	Create a new task belonging to a repository	repository ID, task name
11	Delete a task	task ID
12	Rename a task	task ID, new name
13	View basic content of a task	task ID
14	View either a list of seed patterns along with extracted words, or a list of seed words along with extracting patterns, logs can be listed along with each relation of a pattern and a word, scores will be shown in any case	task ID, a flag to indicate the type of list to be viewed, a flag to indicate the including of logs in the output
15	Add seed words to a task	task ID, a list of words
16	Add seed words from a file to a task	task ID, file name

(continued on next page)

Table 1.2: (continue)

No.	Description	Input
17	Forbid words for a task, these words are considered bad ones and will never be added to the seed set	task ID, a list of words
18	Remove words from a task regardless they are seed or forbidden ones	task ID, a list of words
19	Clear all words in a task	task ID
20	Add seed patterns to a task	task ID, a list of regexes
21	Add seed patterns from a file to a task	task ID, file name
22	Add exceptional patterns to a task, these patterns are considered very good and will be taken first when finding the best ones, words extracted from these are also considered very good and no need to be filtered	task ID, a list of regexes
23	Forbid patterns for a task, these patterns are considered very bad and will never be selected	task ID, a list of regexes
24	Remove patterns from a task regardless they are seed, forbidden, or exceptional ones	task ID, a list of regexes
25	Clear all patterns in a task	task ID
26	Prune the current seed patterns, be careful as some patterns will be removed	task ID
27	Clear all words and patterns in a task	task ID
28	Update parameters for a task	task ID, iteration number, best pattern number, best word number, bad score threshold, pattern generation timeout
29	Run a task	task ID

Total use cases: 29.

Chapter 2

State of the art

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Chapter 3

Technical solution

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Chapter 4

Evaluation

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Chapter 5

Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Bibliography