

USER PROFILING THROUGH ARTIFICIAL IMMUNE  
SYSTEM

By  
Ahmad Raza

SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE  
AT  
BAHRIA UNIVERSITY  
ISLAMABAD  
NOVEMBER 2013

© Copyright by Ahmad Raza, 2014

BAHRIA UNIVERSITY  
DEPARTMENT OF  
COMPUTER SOFTWARE ENGINEERING

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled “**User Profiling Through Artificial Immune System**” by **Ahmad Raza** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: November 2013

Supervisor:

\_\_\_\_\_

Dr. Wasif Afzal

Readers:

\_\_\_\_\_

\_\_\_\_\_

*Dedicated to Sir Wasif Afzal for his continuous efforts and  
guidance and to my parents.*

# Table of Contents

Table of Contents	v
List of Tables	viii
List of Figures	ix
Abstract	x
Acknowledgment	xi
<b>1 Introduction To Web</b>	<b>1</b>
1.1 INTRODUCTION . . . . .	1
<b>2 BACKGROUND AND MOTIVATION</b>	<b>6</b>
2.1 Introduction . . . . .	6
<b>3 Immune System: Natural and Artificial</b>	<b>10</b>
3.1 Introduction . . . . .	10
3.2 Processing of AIS . . . . .	11
3.3 Natural Immune System . . . . .	11
3.3.1 B-Cells . . . . .	12
3.3.2 T-Cells . . . . .	12
3.3.3 Negative Selection . . . . .	13
3.3.4 Clonal Selection Principle . . . . .	13
3.4 Artificial Immune System . . . . .	13
3.4.1 Step 1: Data Extraction . . . . .	14
3.4.2 Step 2: Updating Terms Weights . . . . .	15
3.4.3 Step 3: Remove Unwanted Terms . . . . .	15
3.4.4 Step 4: Add Terms . . . . .	16

3.4.5	Step 5: Reorganization of Profile . . . . .	17
3.5	Clonal Selection . . . . .	17
3.6	Negative Selection . . . . .	18
3.7	Problems in Negative Selection . . . . .	21
3.8	Danger Theory . . . . .	21
<b>4</b>	<b>Applications of Artificial Immune System</b>	<b>22</b>
4.1	AIS in Web Mining . . . . .	22
4.2	Distributive Data . . . . .	22
4.3	Volatile Data . . . . .	23
4.4	Large Data Volume . . . . .	23
4.5	Data Quality . . . . .	23
4.6	Data Type . . . . .	23
4.7	AIS As Intrusion Detection . . . . .	24
4.8	AIS As Data Mining Tool . . . . .	24
<b>5</b>	<b>Research Methodology</b>	<b>25</b>
5.1	Research Methodology We Follow . . . . .	25
5.1.1	Quantitative Research . . . . .	26
5.1.2	Other Methods . . . . .	26
5.2	Need of AIS Based Search . . . . .	27
5.3	Proposed Solution . . . . .	28
<b>6</b>	<b>Enhanced Artificial Immune System (EAIS)</b>	<b>31</b>
6.1	Introduction . . . . .	31
6.2	The Process . . . . .	32
6.3	Search Engine Training Methodology . . . . .	33
6.4	AIS Vs EAIS . . . . .	36
6.5	Comparison With Other Search Engine . . . . .	37
<b>7</b>	<b>Simulation and Testing</b>	<b>40</b>
7.1	Introduction . . . . .	40
7.2	Simulator Overview . . . . .	40
7.3	System Stats . . . . .	42
7.4	Search Mechanism . . . . .	44
7.5	Data Sources . . . . .	45
7.6	System Information . . . . .	45
7.7	Testing . . . . .	46
7.8	Efficiency Comparison . . . . .	47

7.9	User Profile Efficiency . . . . .	48
7.10	Search Query Suggestions . . . . .	49
<b>8</b>	<b>Comparison and Conclusion</b>	<b>51</b>
8.1	Why we use AIS . . . . .	51
8.1.1	Ontology Based User Profiling . . . . .	51
8.2	Research Questions . . . . .	54
8.3	Conclusion . . . . .	55
8.4	Future Work . . . . .	56
	<b>Bibliography</b>	<b>57</b>

# List of Tables

6.1	Traditional AIS Vs Enhanced AIS . . . . .	37
6.2	Search Engine Comparison . . . . .	38
7.1	Search Stats against Term Apple . . . . .	44
7.2	Simulator Stats . . . . .	46
7.3	Simulator Accuracy Test . . . . .	47
7.4	Searching Efficiency Test . . . . .	47
7.5	Algorithm efficiency for Profile Base Search . . . . .	48
8.1	Ontology vs AIS . . . . .	52

# List of Figures

3.1	Adding Terms [1]	17
6.1	Process	34
6.2	Search Engine Training Tree	35
7.1	Username Request Form	41
7.2	Search Engine Main Search Form	42
7.3	Search Keywords Suggestion System	43
7.4	Keyword Suggestion Network	50
8.1	Geography base ontology	53
8.2	Ontology Definition	53
8.3	Semantic Network	56



# Abstract

An Enhanced Artificial Immune system(EAIS) is suggested in this thesis. The proposed system is an enhancement to the traditional Artificial Immune System (AIS) which will monitor user activities during search and will build searching profile for that user. We will use both implicit and explicit ways of declaring user profile and will maintain it according to five steps of AIS. The AIS will be modified by combining some extra parameters to make it more reliable and enable it to give nearest possible search results. This will enhance user experience while searching over Internet. It is also observed that some search engines provide misleading information about well known keywords and terms that cause invalid search results.

To validate the performance of these new models simulation study is carried out by developing a search engine base supporting both AIS and EAIS. The results of simulation are compared with those obtained with Simple Searching algorithm and AIS algorithm. Improved identification and equalization performance of the proposed method have been observed in all cases.

# Acknowledgment

I thank to almighty God who gave me the courage to complete my thesis in time. It will be unjust not to mention the names of those who helped me in getting knowledge and guidance. I am thankful to my teachers, supervisor Dr. Wasif Afzal, Assistant professor at Bahria University Islamabad for his cooperation and continuous suggestions throughout this thesis work. I am also thankful to my parents, brothers and friends and specially thankful to Dr. Shahzad Khalid Associate Professor Bahria University Islamabad for tips on thesis writing and all those who helped me in any way in completing this thesis.