Palestine Polytechnic University
Deanship of Graduate Studies and Scientific Research
Master of Informatics

Social –Driven Agent Based Content Trust Model of Web Resources

Submitted By
Rami Ali Yousef Muhsen

In Partial Fulfillment of the Requirements for the Degree
Master of Informatics

September 2013

The undersigned hereby certify that they have read and recommend to the Deanship of Graduate Studies and Scientific Research at Palestine Polytechnic University the acceptance of a thesis entitled:

Social –Driven Agent Based Content Trust Model for Information Sources

Submitted by **Rami A. Muhsen**

In partial fulfillment of the requirements for the degree of Master in Informatics

Graduate Advisory Committee:

Committee Chair Name, University:

Dr. _____ Date: _____

Signature: _____

Committee Member Name (Supervisor), University:

Dr. Adel Al-Taweel, Palestine Polytechnic University

Signature: _____ Date: _____

Committee Member Name, University:

_____

Signature: _____ Date: _____

External Committee Member Name, University:

_____

Signature: _____ Date: _____

Approved for the Deanship:

Dean of Graduate Studies and
Scientific Research
Palestine Polytechnic University

Signature: _____ Date: _____

Abstract

The number of people using the Internet as a main source of information is constantly growing. And since the content of some of this information could be fraudulent or unreliable, the need for a method to determine its trustworthiness is essential. This research investigates the factors that contribute to trustworthiness of information in web pages, with each factor converted to an agent called *factor analysis agent* that generates a representative value for each factor. We have classified these factors into three bands according to human perception priorities. The research generates a computational trust model based on factors classification as a core of the *Content Trust Agent (CTA)*. The CTA calculates a trust value of information in a given webpage depending on the factor analysis agents' values that communicate between each other and CTA agent to produce our *Multi-Agent System (MAS)*. This work evaluates the outputs of the trust model of our MAS that generate a trust value in a percentage and shows how much the trust value of our model is closed to human perception *(from expert users)*.

ملخص

عدد الأشخاص المستخدمين للإنترنت في العالم كله وفي حياتهم اليومية يزداد باستمرار، باعتبارها المصدر الرئيسي للمعلومات بالنسبة لهم. هذا البحث يحقق في العوامل التي تسهم في مصداقية المعلومات في صفحات الويب باستخدام البحوث التجريبية على أن يحوّل كل واحد من هذه العوامل إلى عميل يسمى "عميل تحليل العوامل" الذي يولد القيمة الممثلة عن كل عامل. قمنا بتصنيف هذه العوامل إلى ثلاثة نطاقات وفقا لأولويات الإدراك البشري. البحث يولد نموذجا حسابياً للثقة على أساس العوامل التي تم تصنيفها باعتبارها جوهر "عميل الثقة بالمحتوى" (CTA). يقوم العميل CTA بحساب قيمة الثقة من المعلومات في صفحة ويب معين معتمداً على قيم "عملاء تحليل العوامل" التي تتواصل فيما بينها وبين عميل CTA لإنتاج نظام متعدد العملاء (MAS). يقوم هذا العمل بتقييم مخرجات نموذج الثقة في MAS والتي تولد قيمة الثقة كنسبة مئوية ويظهر مدى قرب قيمة الثقة من نموذجنا إلى الإدراك البشري (من خلال مستخدمين خبراء).

DECLARATION

I declare that the Master Thesis entitled " *Social –Driven Agent Based Trust Model for Information Sources*", is my own original work, and hereby certify that unless stated, all work contained within this thesis is my own independent research and has not been submitted for the award of any other degree at any institution, except where due acknowledgment is made in the text.

Rami Ali Muhsen

Signature: _____ Date: _____

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for the master degree in Informatics at Palestine Polytechnic University, I agree that the library shall make it available to borrowers under rules of the library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of the source is made.

Permission for extensive quotation from, reproduction, or publication of this thesis may be granted by my main supervisor, or in his absence, by the Dean of Graduate Studies and Scientific Research when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Rami Ali Muhsen

Signature: _____ Date: _____

DEDICATION

I dedicate this work to my parents, my wife and our beautiful daughter, without whose support, patience and encouragement, the work wouldn't have been completed successfully.

ACKNOWLEDGMENT

The author would like to thank the supervisor in this thesis Dr. Adel Al-Taweel, for his efforts and hosting this research. He taught me the methodology of research and offered me great freedom on choosing my favorite research topic. He also guided me with his valuable comments during this work. My thanks also go for Palestine Polytechnic University (PPU) lecturers and students for their valuable assistance and support during empirical experiments in this research. I thank the McAfee trusted source team, ranking.com Team and SiteOpsys researcher and developer team for making the data in this work available. Our thanks connected also to the examining committee, all of my colleagues and relatives for their support.

Abbreviations

Abbreviations	Full term
1. **ABM**	Agent Based Model
2. **ws**	Website
3. **PHP**	Personal Home Page
4. **IP**	Internet Protocol
5. **URL**	Uniform Resource Locator
6. **SQL**	Structured Query Language
7. f	Factor
8. f_n^j	Number Of Times A Factor n Ranked As j
9. w_{f_1}	Weight For Factor1
10. v_f	The Value Returned From Factor Agent
11. ***wp_url***	Webpage URL
12. **SHA**	Secure Hash Algorithm
13. **NIST**	National Institute of Standards and Technology
14. **HTTP**	Hyper Text Transfer Protocol
15. **XML**	Extensible Markup Language
16. **HTML**	Hyper Text Markup Language
17. **SSXML**	Style Store XML
18. **CSS**	Cascade Style Sheet
19. **API**	Application Programming Interface
20. **J-AAPI**	Java Aglet API
21. **JDK**	Java Development Kit
22. **ASDK**	Agent Software Development Kit
23. **CTA**	Content Trust Agent
24. **MAS**	Multi-Agent System
25. **PPU**	Palestine Polytechnic University
26. $\mathbf{Ws_ng_m}$	Website *n* Group *m*
27. **HR**	High Rational
28. **MR**	Medium Rational
29. **LR**	Low Rational

TABLE OF CONTENT

TITLE	PAGE

ABSTRACT ..... **III**

ملخص ..... **IV**

TABLE OF CONTENT ..... **X**

LIST OF FIGURES ..... **XIII**

LIST OF TABLES ..... **XV**

LIST OF ALGORITHMS ..... **XVI**

CHAPTER 1: INTRODUCTION **2**

1.1 INTRODUCTION 2
1.1 OVERVIEW 4
1.2 MOTIVATION 5
1.3 PROBLEM STATEMENT 5
1.4 RESEARCH METHODOLOGY 5
1.5 THESIS OUTLINES 6

CHAPTER 2: LITERATURE REVIEW **9**

2.1 INTRODUCTION 9
2.2 TRUST MODELS 9
2.3 TRUST MODELS IN MULTI-AGENT SYSTEMS: 11
2.4 EMPIRICAL RESEARCH: 12

CHAPTER 3: EMPIRICAL EXPERIMENT FOR TRUST FACTORS IDENTIFICATION AND ANALYSIS **15**

3.1 INTRODUCTION 15
3.2 CONTENT TRUST FACTORS COLLECTION 15
3.3 EMPIRICAL EXPERIMENT DESIGN 16
3.4 EMPIRICAL EXPERIMENT RESULTS ANALYSIS 18
3.4.1 Factors Identification and Filtration *18*
3.4.2 Factors Prioritization *22*

CHAPTER 4: FACTOR ANALYSIS AGENTS **28**

4.1 INTRODUCTION 28
4.2 FACTOR ANALYSIS AND CTA AGENTS ENVIRONMENT 28
4.3 FACTOR ANALYSIS AGENTS IMPLEMENTATION 30
4.3.1 Authority Factor Agent *30*
4.3.2 Specialization and Security Factors Agent *30*
4.3.3 Popularity and Number Of Visitors Factors Agent *31*
4.3.4 Age and Update Factor Agent *32*
4.3.5 Methodology Of Writing Factor *33*
4.3.6 Links Inconsistency Agent: *36*
4.3.7 Appearance Of Website Agent: *36*

CHAPTER 5: CONTENT TRUST AGENT (CTA) MODEL **39**

5.1 INTRODUCTION 39
5.2 PROPOSED MODEL 39
5.3 FACTORS WEIGHTS MATHEMATICAL FORMULA 39
5.4 FACTORS CLASSIFICATION 41
5.5 CONTENT TRUST MODEL ALGORITHM 43
5.6 MODEL SELECTED EXAMPLES 45

CHAPTER 6: CTA AGENT IMPLEMENTATION 49

6.1 INTRODUCTION 49
6.2 CTA AGENT ALGORITHM AND PROGRAMMING 49
6.2.1 CTA Agent Algorithm 49
6.2.2 CTA Agent Programming 52
6.3 AGENTS RUNNING 52
6.3.1 Introduction To Aglet 52
6.3.2 Running Aglet Framework (Tahiti) 52
6.3.3 Running Content Trust Agent On Tahiti 54

CHAPTER 7: MODEL EVALUATION 58

7.1 INTRODUCTION 58
7.2 EVALUATION METHOD 58
- *Confidence Interval: 59*
- *Hotelling's T^2: 60*
- *Evaluation of Model Design: 60*

7.3 EXPERIMENT RESULTS ANALYSIS 61
7.3.1 Factor Analysis Agents Results 61
7.3.2 CTA agent and Respondents Information Analysis 63
- *Confidence Interval Analysis: 63*

7.3.3 Analyze Demographic Information 68
- *Gender Analysis: 68*
- *Job Experience Analysis: 72*
- *Academic Education Analysis: 73*
- *Webpages Article Exceedance: 74*

7.3.4 Factor Analysis Agents Collaboration Effects 75
- *MR and LR Factors Effects On HR: 75*
- *Factor Analysis Agents Dependences: 76*

7.3.5 Case Study: Medical Webpages Content 77
7.4 RESULTS 80

CHAPTER 8: CONCLUSION AND FUTURE WORK 83

8.1 CONCLUSIONS 83
8.2 LIMITATION 84
8.3 FUTURE WORK 85

REFERENCES 87

APPENDIX A 93

APPENDIX B 98

APPENDIX C 100

- VALIDATION DESIGN: 101
- VALIDATION RESULTS: 101
- VALIDATION CODE: 102
- SELECTED VALIDATION RESULTS: 105

APPENDIX D 113

APPENDIX E 122

APPENDIX F 126

- *Hotelling's T^2:..... 127*
  - *Hotelling's T^2 Analysis:* 127
  - *Hotelling's T^2 Analysis Excluding Missed Webpage For Group1:* 130

List of Figures

Figure	Page

Figure 1.1: Internet Is Top Source for People Answers 2

Figure 3.1: Start PHP page of the Questionnaire. 17
Figure 3.2: PHP page of the Questionnaire. 18
Figure 3.3: The percent of factors as a results from the questionnaire. 20
Figure 3.4: Categorize Resources Content Trust Percent From Experiment. 21
Figure 3.5: Percent of factors mentioned for each website category. 26

Figure 4.1: CTA and Factor Agents Environment. 28
Figure 4.2: Flowchart for Preparation of Factor Analysis Agent to be Dispatched to agents Environment. 29
Figure 4.3: LanguageTool Application. 33
Figure 4.4: LanguageTool Options. 34
Figure 4.5: CSS Merging Process. 36
Figure 4.6: Architecture of Style Store XML File. 37

Figure 5.1:Factors Weights vs. Weight Percent Features 41
Figure 5.2:Factors classification according to factors count and weight features. 43

Figure 6.1: a) CTA agent send Future-type(asynchronous) to LR factor agent exploits wait time by sending as it can Now-type messages to HR factor agents . b) shows that a three Now-type messages to HR factor agents are exploited when LR factor agent respond. 50
Figure 6.2 .:Tahiti Server Running. 54
Figure 6.3:Agent Architecture In Aglet. 54
Figure 6.4 .: Create A New Agent On Tahiti Server. 55
Figure 6.5: Create A New Agent Window. 55
Figure 6.6: Content Trust Agent Running In Tahiti 55
Figure 6.7 .:Content Trust Agent Results In CMD. 56

Figure 7.1: 95% confidence interval for μ. 60
Figure 7.2: T-Distribution Table For Each Number of Population 63
Figure 7.3: Respondents average answers vs. CTA agent answers within $[\mu - t_{n-1}\frac{\sigma}{\sqrt{n}}, \mu + t_{n-1}\frac{\sigma}{\sqrt{n}}]$. (group 1) 64
Figure 7.4: Respondents average answers vs. CTA agents answers within $[\mu - t_{n-1}\frac{\sigma}{\sqrt{n}}, \mu + t_{n-1}\frac{\sigma}{\sqrt{n}}]$. (group 2) 65
Figure 7.5: Respondents Gender Distribution. 68
Figure 7.6: CTA agent answer compared to average male respondents answers and confidence interval /Group1. 69
Figure 7.7: CTA agent answer compared to average female respondents answers and confidence interval /group1. 70

Figure 7. 8: CTA agent answer compared to average male respondents answers and confidence interval /Group2. 70
Figure 7. 9: CTA agent answer compared to average female respondents answers and confidence interval /Group2. 71
Figure 7. 10: Respondents Job Experience Distribution. 72
Figure 7.11: The accuracy results for group 1 (a): for 1-5 years (b):More than 5 years job experience. And for group 2 (c): for 1-5 years (d):More than 5 years job experience. 73
Figure 7. 12: Respondents Majors Distribution. 74
Figure 7. 13: (a) The results and accuracy of engineering respondents. (b) The results and accuracy of management/ financial/accounting respondents 74
Figure 7. 14: Effect of MR and LR factors on HR factors for group1 75
Figure 7. 15: Effect of MR and LR factors on HR factors for group2 75
Figure 7. 16: CTA agent answer compared to average doctors answers on medical webpages. ..78

Figure F.1 : upper percentage points of the F distribution 129

List of Tables

Table	Page

Table 2.1: The Differences Between Quantitative and Qualitative research methods 13

Table 3.1: The Main Factors Extracted From Experiment. 19
Table 3.2:Resources Content Trust Percent From Experiment. 21
Table 3.3:Factors Order For Social Sites Category. 22
Table 3.4:Factors Order For Medical Sites Category. 23
Table 3.5:Factors Order For News Sites Category. 24
Table 3.6:Factors Order For Scientific Sites Category. 24
Table 3.7:Factors Order For Governmental Sites Category. 25

Table 4.1:Reputation Scores. 31

Table 6.1:The message passing types used by CTA agent for each classification of factor analysis agents. 49

Table 7.1:Factor analysis agents returned values for first group. 62
Table 7.2: Factor analysis agents returned values for second group. 62
Table 7.3: Respondents results compared to CTA agent calculation (First Group). 66
Table 7.4: Respondents results compared to CTA agent calculation (Second Group). 67
Table 7.5 :Missed webpages summary according to gender. 71
Table 7.6: Demographic results regarding experience. 72
Table 7.7: Missed webpages summary according to job experience. 73
Table 7.8: Demographic results regarding the majors. 74
Table 7.9: Respondents vs. CTA agent results for HR, MR and LR (First Group). 75
Table 7.10: Respondents vs. CTA agent results for HR, MR and LR (Second Group). 76
Table 7.11: The Effect Of Locality On Popularity Factor 76
Table 7.12: The Effect Of Security On Specialization Factor 77
Table 7.13: Respondents results compared to CTA agent calculation (Medical Article). 79

Table C.1: A Sample of Model Validation. 101

Table F.1: $(X - \mu_0)'$Vector Values From Respondents Mean Vector and Agent Answers Vector/Group1 128
Table F.2: $(\boldsymbol{X}-\boldsymbol{\mu 0})'$Vector Values From Respondents Mean Vector and Agent Answers Vector/Group2 128
Table F.3: S^{-1}For Group1 Dataset 131
Table F.4: S^{-1}FOR GROUP2 DATASET 131
Table F.5: $(X - \mu_0)'$Vector Values After Excluding *Ws10g1, Ws12g1 and Ws13g1*/Group1 132
Table F.6: S^{-1}FOR GROUP1DATASETAFTER EXCLUDING/ Group1 132

List of Algorithms

Algorithm	Page

Algorithm 4.1 .:Online Capture Age and Update Factor 32
Algorithm 4.2 .:Online Check Grammar Of Webpage Content 34
Algorithm 4.3 .:Online Check Spelling Of Webpage Content 35

Algorithm 5.1 .:Model Algorithm 44

Chapter One

Introduction

Chapter 1: Introduction

In this chapter we present a brief overview of trust, information sources, information sources trust in web. Motivation of thesis is also introduced, problem statement, proposed solution and research methodology.

1.1 Introduction

As illustrated in survey [39], most people sought information through a variety of resources, including Internet, professional advisers, "friends and family" and others. The survey was conducted through a sample of 2796 people, who were asked what type of information sources they used when they addressed a certain problem from their point of view, As the survey results in Fig. 1.1 show, "internet" is the top source of information.

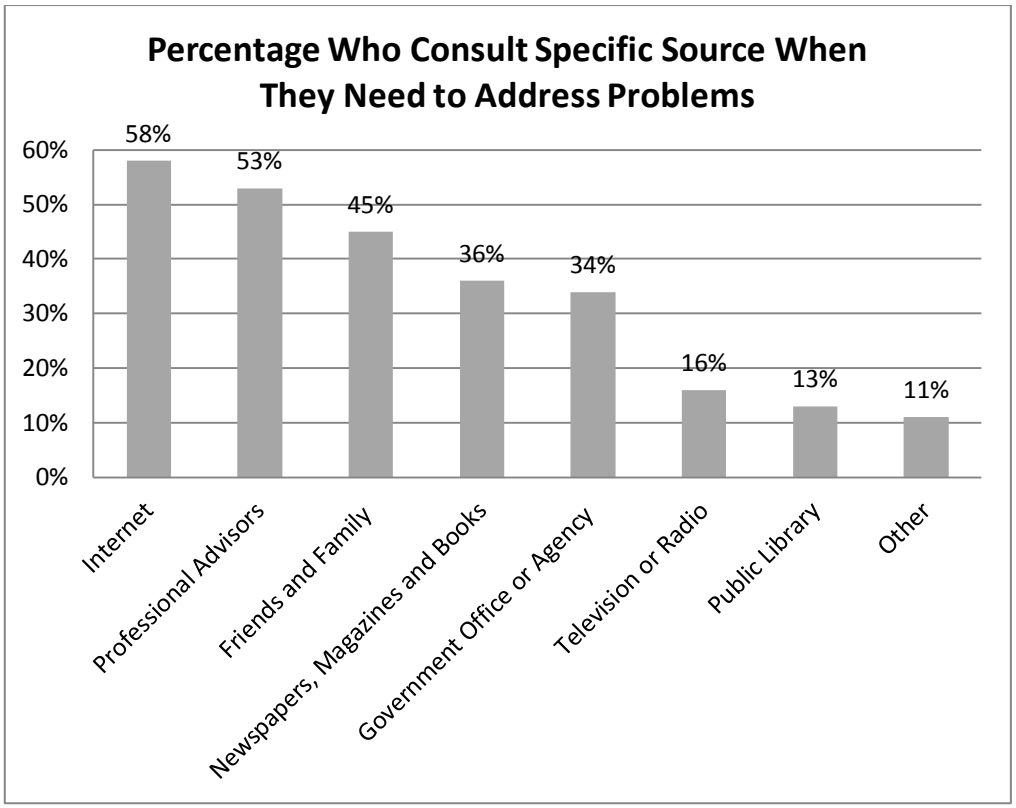


FIGURE 1. 1: INTERNET IS TOP SOURCE FOR PEOPLE ANSWERS

"The accuracy of internet information is a central issue that needs to be studied rather carefully in order to avoid web hoaxes and misinformation. Misinformation on the internet is, and will always be, a problem. A key problem of the internet is the fact that nearly anyone can publish data on the worldwide web which creates an environment of freedom that make hard control on the quality. The internet users will resorted to more confident resources to perform the necessary check about information since the quality is absent"[37].

Misinformation does not necessarily get understood as "wrong" information; since most of them could appear as opinions or biased rather than facts[37]. For example, family planning might solve a lot of economic issues in some countries, hence the information will written in context of encouragement reflecting governmental point of view in this regards. In contrast, antithetical information might appear in other countries against family planning given the proper context [37].

The web hoaxes and misinformation affect many aspects of our daily life such as medical, educational and businesses spheres[38][40]. The work in [38] discusses the effect of inaccurate health information on a group of 34 students from the Science Magnet High School in Houston, Texas. "By asking them to search for the terms "vaccine safety" and "vaccine danger" using Google and then answer questions regarding the accuracy of the health information on the returned sites"[38]. The study confirmed that students might gain incorrect information about medical topics, and they needs to be more carefully in medical web content. Another work in [40] studying the quality of information that exists on the internet about "depression" by making a search about the phrase "depression and treatment" on ten major internet search engines. "The researchers then examined the first 20 sites generated by each engine and analyzed the returned results, they concluded that the quality of information was quite low, prompting us to think how we can control this information to be in the right direction"[40].

One potential solution to control the use of misinformation from the internet is using the notation of *"Trust"*. To illustrate how we can evaluate trust in our research, let's take a scenario that describe a particular situation. Now, imagine a salesman knocking on your door. You've never met him before, but he wears a suit belonging to a company, and he has a bag containing some merchandise belonging to the same company and wears a suit. He speaks with you in a tactful manner, asking if you desire to buy some products and shows you a certificate as accredited salesman for the company. Do you believe that he's a salesman really? Do you believe that his

merchandise is good? The answer is mostly like to be yes. Now suppose the salesman was in jeans and a T-shirt, and carries with him a bad looking bag with different types of merchandises and without any company labels. He speaks rudely and incessantly, you would probably take a lot more time accepting his products and asking lots of questions, such as who the man is, where he's from, and so on. In the first instance, you *trust* the first salesman because of what he is, or at least, what you can see he is. In the second, you will have a serious *trust* problem with the man in jeans and the way of his speaking, seeking for much reassurance factors.

We can notice from the last example that one way of evaluating trust is attained through social factors reflecting human perception about "Trust" in something in life and this represent the path of this research to find potential solution of discovering and guarding against misinformation on web.

1.1 Overview

The "*trust*" can be defined as the "firm belief in the reliability, truth, or ability of someone or something" [43], another historical? Based definition of trust from Mui et al said: "Trust is a subjective expectation an agent has about another's future behavior based on the history of their encounters"[44]. The *web content trust* is a way that could be used in order to check whether a web content from a specific information source is credible or not.

An *information source* is a "source of information for somebody, i.e. anything that might inform a person about something or provide knowledge to somebody" [31]. In web, information sources come from many websites or web pages to many users over the internet by reading articles, news and any information they are concerned with. *Web information sources trust* means the study of articles and information trust in web pages over Internet from readers' point of view.

"Modeling" refers to "the process of generating a model as a conceptual representation of some phenomenon"[45], while *trust model* is a way to represent or embody the trust to serve someone seeking it ideally. *Web information sources trust model* is a representation of trust judge for web-page articles and information validity over internet and how much it is closely related to readers judge.

1.2 Motivation

The Internet is a huge repository of information that is available for every user; this availability leads people to rely on the Internet even for the simplest tasks. The absence of any validation and accuracy on the web content render information fraudulent. This leads us to think how we can make sure that we got accurate information from any resources in the Internet. For example, if users wanted to search for a piece of information over the Internet (e.g. news), they would search through using a search engine, then all related websites would be returned without any mechanism to ensure whether this returned information from these websites is accurate or not.

1.3 Problem Statement

We need to find a method to tag a website with a value of trust. In real life, this will be an intuitive measure for any user to trust a piece of content on a website. In this research we are creating computational measure of accuracy of information based on scientific method. So we will study factors that affect trust and develop a trust model based on agent approach to derive trust measure. This research will primarily focus on web content from web 1.0 sources. Web 2.0 web sources and technologies are outside the scope of this research.

1.4 Research Methodology

In order to calculate the trust measure of web content from human perception, our research follows four main steps:

- *Factors Identification and Classification:*

For this step, in this research, factors were identified through two mechanisms: factors already identified in the relevant literature review and an empirical study. An empirical study was designed to identify the factors from a human perception point of view that are percieved (by the participants) to affect trust in websites' content, We achieve this by performing experiment on a set of people divided into groups (groups were selected carefully) to view selected websites (websites containing the factors implicitly), then the results will be analyzed to select the most common factors and their significance (see section 3.2 in chapter 3). The numbers of factors are huge and each of them has different effects on trust model, so these factors are filtered and prioritized empirically depending on how much a factor was approved by different people and websites categories (see section 3.3.2 in chapter 3).

These factors are classified then into three categories based on *factor count* and *factor weight* features (see section 5.2 in chapter 5). The three categories are: *low rational, medium rational* and *high rational factors*. Each category contains a set of factors that have the same level of importance (see section 5.4 in chapter 5).

- *Factor Analysis Agents Evaluation:*

Agents, called factor analysis agents, are developed to calculate each content trust factor. The *Content Trust Agent* (CTA), acting as a coordinator, collaborates with all respective factor analysis agents for a given web source (or URL) to derive an overall trust value for the source in question. Each factor analysis agent returns a value from 0 to 9 representing how much confidence this factor appears on this webpage article URL *(0 Not at all, 9 Completely appears)*

- *CTA Agent and Trust Model:*

The *CTA* was built to deal with other factor analysis agents by passing a webpage article URL to them, then it waits until it has all responding values from factor analysis agents. After receiving all values, it starts its calculation of content trust percent *(from 0% to 100%)* using our *trust model* (see section 5.4 in chapter 5).

- *MAS Implementation and Evaluation:*

The agents *(including CTA and factor analysis agents)* running on aglet framework using *Tahiti* server to be prepared for communication, learning and provide the trust value from CTA agent when it is requested. The developed trust model, CTA agent and 8 factor analysis agents are evaluated using an empirical study of 40 expert users utilising 30 information sources *(divided into 2 groups with 20 experts and 15 information sources each)*. The evaluation povides an indication of the trust judgment between our multi-agent system and expert answers, thus is used as a benchmark to note the error ration of our trust model.

1.5 Thesis Outlines

The organization of this thesis can be summarized as the following:

1. Chapter one includes a brief **Background and Overview** of trust, information sources and information sources trust in web. Motivation of thesis, problem statement, proposed solution and research methodology.

2. Chapter two presents an extensive **Literature Review** covering all related works, in addition to filtration of works pertaining to our research.
3. Chapter three introduces the **Empirical Experiment** to identify content trust factors that include: literature review trust factors elicitation, empirical experiment design and empirical experiment results analysis.
4. Chapter four presents the methodology of preparing **Factor Analysis Agents** and how these agents extract their information and generate their own knowledge (*factors databases and services*).
5. Chapter five proposes a new mathematical model to generate trust value, factors weights calculation.
6. Chapter six presents agent algorithm and programming, agent running
7. Chapter seven presents agent evaluation through experimental design for users and compares the results from user to results from our agent.
8. Chapter eight presents a comprehensive conclusions and future work.

Chapter Two

Literature Review

Chapter 2: Literature Review

2.1 Introduction

In this chapter we present a *literature review* about all related works especially in trust models and multi-agent system trust models, besides filtration of most relevant ones to our research. We introduce the *empirical research* as a mean to clarify this term to the reader in addition to their types.

2.2 Trust Models

As we mentioned in the previous chapter, modeling refers to the process of generating a model as a conceptual representation of some phenomena, while *trust model* is a way to represent or embody the trust to serve someone seeking it ideally. *Web information sources trust model* is a representation of trust judgment for webpage articles and information validity over Internet and how much it is closely related to user judgment. We will introduce all related works in trust models and discuss how these models address the trust variable.

The Ph.D. thesis in [30] by Marsh is one of the best works that address computational models of trust rather extensively. His purpose was to process "an imperfect understanding, a plethora of definitions, and informal use in the literature and in everyday life" with respect to trust. He proposed a set of (subjectively-based set) variables, and a way to combine them to arrive at one continuous value of trust in the range [−1, 1]. While the intuitive explanation of this may range from complete distrust to full trust, Marsh actually argues against these meanings at the extremes, arguing that neither full trust nor distrust is actually possible. Marsh identified three types of trust: basic, overall contexts; general, between two people and all their con-texts occurring together; and situational, between two people in a specific context. In addition to context, Marsh also identified time as being relevant to each of the variables used to comprise trust. Authors who cite Marsh frequently use a simplification of his work (e.g., trust is a continuous value, and its composition is not of concern) or do not follow his model due to the difficulty of finding values for some variables used to compute trust (e.g., importance, utility, competence, risk, etc.).

The work in [13] is closest to our research and represents a good start in our endeavor. The authors studied *content trust factors* that users consider in deciding whether to trust the content provided by a web resource, which this step represents as the base of our work. We used the 19 factors resulting from [13] (in addition to brain storming and a lot of search through the web) to select webpage articles during empirical experiment design. The experiment seeks to achieve two things: First, determine which one of these factors is related to human perception. Second, ranking these factors according to their significance and relevance to human perception. The authors used associations concept (e.g., Referencing and Authority) to transfer trust from entities to resources preferring this way over using any trust heuristics (e.g., Appearance, Age and update) The authors mention the factors of content trust without any clarification as to how these factors are precisely identified? And which one of these factors is identified with and close to human perception?

A model is proposed based on beliefs and their credibility for making trust decisions about sources, differentiating between internal and external attributes affecting trust in a source [32]. In our work the model based on factors, which are identified socially, represents the inputs to *content trust agent* CTA, the factors are not confined to four inputs but include all variables affecting trust associated with their effect rate. The authors note that the composition of inputs to a trust decision affects the outcome of the decision, and thus the decision itself cannot be characterized by a final probability. This observation might be restated that the inputs together form part of the context in which trust is being determined. Also acknowledged is that "attribution of trust is a very complex task", a problem that is exemplified on the Web, as the sources behind information are not always clear or correct. Specifically for trust in information sources, four types of inputs to a trust decision are given: direct experience, categorization (generalization to or from something known), reasoning (application of common sense or rules to verify truth), and reputation.

In [25], the authors use the notion of content trust for spam detection using *evidence*-based system. The word "*evidence*" here is relatively equivalent to "*factor"* in our work, since they concentrate on traditional text feature attributes (e.g., keyword stuffing, features of the host component of a URL, Excessive replication of content, …etc.) and six information quality (e.g., Currency, Availability, Information-to-noise ratio, Authority, … etc.). The authors' work does not show on what scientific base they select these evidences, and what the users opinions are on these evidences. Are they representing a good measure to detect a spam webpage? Besides, the authors ignored if the possibility of more *evidences (factors)* that could have a critical effect on

their system. So the evidence technique is helpful to improve the precision of our agent in addition to trust factor influences.

Another work [11] presents a method to predict Wikipedia articles trustworthiness based on computational trust techniques and a deep domain-specific analysis. The trust computation they build based on Wiki propositions (i.e. authorship problems, accuracy, …etc.), is similar to our factors that we want to deeply analyze, but we would like to include several other websites beyond just Wiki pages. One of the strong attributes of Wiki pages is the speed at which can be updated, which enables us to deal more adequately with these pages in our model.

2.3 Trust Models in Multi-Agent Systems:

Our work doesn't research multi-agent systems and is primarily focusing on trust models since we are using multi-agents in our work for to collaborate to solve more complex problems that can't be solved by one agent. Multi agents can be used to solve problems that it is difficult or impossible for a monolithic or a single agent system to resolve [41]. This represents the environment of how CTA agent and factor analysis agents collaborate and communicate to achieve the content trust percent. We will introduce some background about agents, multi-agent systems and related works of trust models in multi-agent system.

The definition of "*Agent*" tends to agree on several features common to most agents like (Autonomy, Heterogeneity, Active, Pro-active / goal-directed, Reactive / Perceptive, Interactive / Communicative, Mobility and Adaptation / Learning) [9]. A specific type of system that is composed of multiple intelligent agents that interact with each other to achieve certain objectives is called *agent-based model (ABM) or multi-agent system (MAS)* is. These systems can be used to solve problems that it is difficult or impossible for a monolithic or a single agent system to resolve. However some define MAS as a field that concerned with building of an intelligent software program based on the concept of "Agent", besides "it's a class of computational models for simulating the actions and interactions of autonomous agents (both individual or collective entities such as organizations or groups) with a view to assessing their effects on the system as a whole. It combines elements of game theory, complex systems, emergence, computational sociology, multi-agent systems, and evolutionary programming. [5][41].

Our work is modeled as multi-agent systems that are composed of autonomous factor analysis agents that interact with CTA agent using particular mechanisms and protocols. The agent's research community has developed a number of models of interactions including coordination

[14], [15], [24] collaboration [34] and negotiation [35], [36]. However, in our work we focused on collaboration between factor analysis agents and CTA agents, while CTA agents societies and multi-factor analysis agent could be studied in future works.

2.4 Empirical Research:

Although our research does not include any research development into empirical research, however for the sake of clarity, we introduce empirical research for the benefit of the reader. *Empirical research* methods are "a class of research methods in which *empirical observations* or data are collected in order to answer particular research questions. While primarily used in academic research, they can also be useful in answering practical questions" [33].

As noted in [33], "empirical research usually begins with a theory that requires a researcher to explain and/or predict what happens in the real world. The aim of empirical research is to test the theory and refine it, while in some cases the research is concerned with making these theories, which makes the research harder and risky" [33]. From the empirical experiment definition, we must first define research question. For example: What are the critical successes factors affecting web content trust? To be empirically tested, the next step is to transform the research question into a theoretical model that could be consisting of theoretical constructs (latent variables), causal relationships and measures (observed variables). The analysis of the literature represents the core of theoretical model developing. The theoretical model forms the basis both for collecting and analyzing data, and may be modified depending on the evolution of the research. A hypothesis defines an expected relationship between variables (based on causal relationships in the theoretical model), which can be empirically tested [33].

Empirical research methods could be: *Quantitative research methods* or *Qualitative research methods[33]*. The differences between qualitative and Quantitative methods is summarised in table 2.1[33]:

TABLE 2. 1: THE DIFFERENCES BETWEEN QUANTITATIVE AND QUALITATIVE RESEARCH METHODS

Qualitative Research	***Quantitative Research***
Collect qualitative data (data in the form of text, images, sounds) drawn from observations, interviews and documentary evidence, and analyze it using qualitative data analysis methods.	Collect numerical data (data in the form of numbers) and analyze it using statistical methods to identify patterns and relationships in the data.
More appropriate in the early stages of research (exploratory research) and for theory building.	More appropriate when theory is well developed, and for purposes of theory testing and refinement.
Qualitative methods could be: Case study or Action Research	Quantitative methods could be: Experiment, Survey or Historical data
Tend to be applied more easily in real world settings	Tend to result in more convincing scientific evidence, they are generally more difficult to apply in a real world context (field setting).

(Source:[33])

In our work we follow *quantitative methods* where we run a set of experiments and surveys to identify instinctively web content trust factors from respondents including ask questions (face to face interview, telephone, mail, Internet) we collect a historical data to look for patterns in historical data (e.g. content trust factors patterns) while quantitative methods tend to result in more convincing scientific evidence.

Chapter Three

Empirical Experiment for Trust Factors Identification and Analysis

Chapter 3: Empirical Experiment For Trust Factors Identification and Analysis

3.1 Introduction

In this chapter we will introduce the method that we follow to collect content trust factors and the methodology of designing our empirical experiment. After that we will show the result of experiment and how we will filter and prioritize them according to their importance from respondents point of view.

3.2 Content Trust Factors Collection

This research made a deep investigation into all factors that influence trust of resource content based on: Literature Review as in [13], [25], [11], Brain Storming and Web Search as in [42], [49], [47], [48] and [46]. Most of the factors were concerned with trusting resources itself not their content. However we collected these factors in order to find any relation between "resource trust" and its "content." After applying these techniques we got over 25 factors, then we summarized them into 17 most previesly discussed factors as following:

1) **Referencing**. If the information is referenced well, it will convince people how much these information is precise and trustworthiness.
2) **Popularity**. The more popular resources are more trusted in its contents.
3) **Reputation.** How many users are referring to a resource, which provide reputation information.
4) **Repeating.** Repeating information through multiple resources provide some amount of trust to judge the information in these resources.
5) **Authority.** If the resource is authorized, then the trust of its content increasing, for instance people are more trusted in known news source than anonymous personal page.
6) **User Experience.** The knowledge of a user with resource information content, which provides a judgment based on his/her experience.
7) **Neutralization/Bias.** A biased source leads to fraud information. For instance if we have information from political party resource, this leads to minimizing trust of these information since they are not written in neutral point of view.
8) **Specialization.** A specialized resource provides user customized information that increases the trust feeling toward this information.

9) **Age and Update.** Frequent updating of content, at least once every few weeks, and more often, once a week or more.
10) **Appearance.** A good design and professional look of resources makes the user feeling more trust in its content.
11) **Methodology of Writing.** This includes grammar and spelling of the content if they are correct or not, and their influences in trust.
12) **Security.** If a resource classified as minimal risk resource, then it will be trusted in its content than high-risk resources.
13) **Seriousness.** How much this resource is serious in its information, if seriousness exists in information, it leads to a good trust indication.
14) **Multimedia Supporting.** This means that the information that is supported by images, videos, …and etc. end to be more trusted in its content.
15) **Locality.** The user that reads information from a resource that is near to where he lives increases the trust of information.
16) **Proximity.** How much this information is approximate to original source. If they are nearest, the trust value increase.
17) **Feedback.** The ability of users to provide the resource with their own knowledge about the information. This makes a refinement on information.

The factors identified in this stage will be used as a base of empirical study.

3.3 Empirical Experiment Design

The empirical experiment used a survey aimed to identify the resources content trust factors from the respondents without any bias. Providing or pre-empting pariticipants with a list of the factors of interest may end up introducing these in their thinking process and thus introducing a bias to their selection. However the aim was to try to let the participants identify the factors themselves with no or minimal bias. In doing so, we can use the pariticipants identification and prioritisation to classify and attach a priority to each factor based on how they percieved the importance of each as a factor of trust. To ensure the selected web sites included the factors of interest for our model, this was achieved by *manually* pre-selecting websites information or articles. Manually means that the selected the webpages hold a content, design and information characteristics that must have implicitly some level of factors of interest we identified in the previous section. For example, for the appearance factor, several sites were selected to have varied levels of appearance between professionally designed interfaces with hormonous “feel and look” to badly

designed interfaces (according to in the opinion of the study designer). During the study, we devised an automated survey mechanism to monitor and allow the respondents instinctively respond to determine factors priority (*which factors become first and which become last*).

The study included two groups of respondents (groups will be created carefully) *10 pariticpants* in each. Each participant in each group was asked to view a set of websites (*25 websites*) randomly selected to each at the time of starting their experiment. The 25 websites for each group contain different set of factors, that were implicitly pre-selected. The selected websites are divided into five categories: social, medical, news, scientific, and governmental sites. The experiment executed in Palestine Polytechnic University laboratories using PHP page (See figure 3.1, figure 3.2) designed especially for this experiment asking the following questions after user read a 5 minute article in each website:

- Do you trust this information? (YES, NO)
- What factors make you trust in it? (At least 3 Factors.)
- Prioritize most three factors?
- Any Notes?

FIGURE 3. 1:START PHP PAGE OF THE QUESTIONNAIRE.

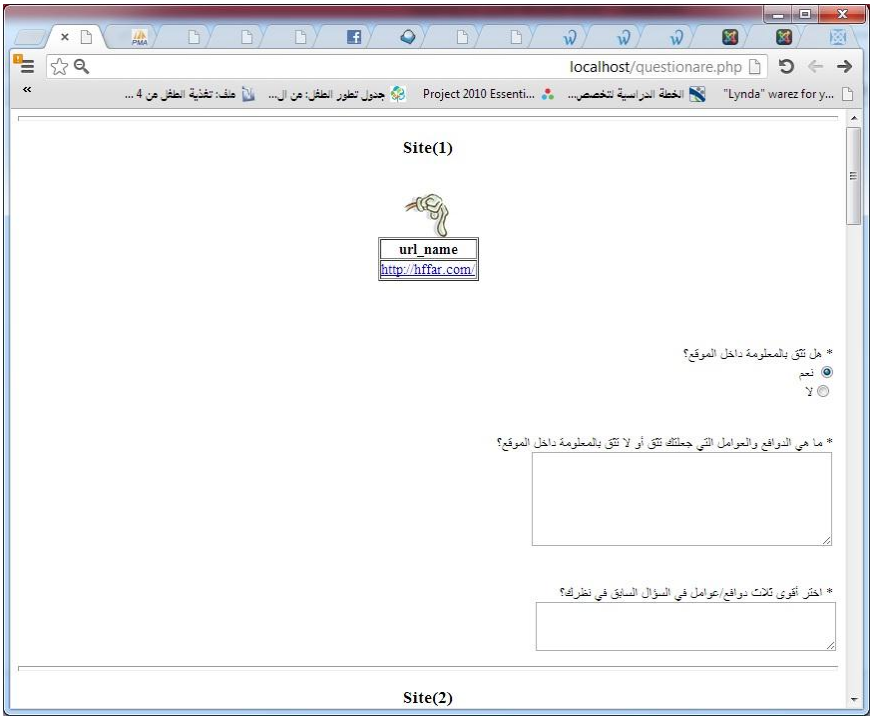


FIGURE 3. 2:PHP PAGE OF THE QUESTIONNAIRE.

At the starting page of our questionnaire, the software begins by storing respondents IP addresses, then asking the respondent about his/her gender, both IP address and gender stored in *PERSON table*. After that for each respondent it generates a random order of website articles from stored MySQL database *URL table* and preserving the order and answers in *ANSWER table*. The questionnaire page is just concerned with viewing the articles for the respondent as ordered previously and stores the answers of the respondents.

3.4 Empirical Experiment Results Analysis

3.4.1 Factors Identification and Filtration

We will deal with experiment results through two stages: *factors identification and filtration* and *factors prioritization.* The number of factors resulting from this experiment is huge which could reach maximum to: 20x25x3= *1500 factors* (20 persons, 25 websites and 3 answers for each respondents) and minimum count of factors could be: 20x25x1= *500 factors* (20 persons, 25 websites and 1 answer for each respondents), the actual number of factors we have in this experiment are *861 factors*.

This large number of factors needs a mechanism to *filter* these factors empirically like:

- Merging Synonymous Phrases.

- Eliminating Non Reasonable Answers.
- Factors can't be implemented.
- Left it Blank

The final factors are shown in table 3.1, which represents 21 factors after applying filtration. The percentage of how much each factor is mentioned by respondents is shown in figure 3.3. We can notice that the referencing factor represents the most effective one on content trust from the respondents point of view, however appearance of website comes into the second place and specialization and authority come in third and fourth places respectively. The other factors as showed in figure 3.3 come to decrease on its percent of trust ending with repeating factor.

TABLE 3. 1: THE MAIN FACTORS IDENTIFIED FROM EXPERIMENT.

Factor	Count	Factor Percent			
Referencing	111	14.07%			
Appearance of Website	104	13.18%			
Authority	91	11.53%			
Specialization	90	11.41%			
The Methodology of Writing	57	7.22%			
Seriously	53	6.72%			
Age and Update	52	6.59%			
Popularity	41	5.20%			
Multimedia Supporting	31	3.93%			
User Expertise	31	3.93%			
Proximity	23	2.92%			
Locality	22	2.79%			
Links Inconsistency	18	2.28%			
Evaluating and Comments	17	2.15%			
Neutralization	16	2.03%			
Contact	14	1.77%			
Constrains	6	0.76%			
Feedback	5	0.63%			
Security	3	0.38%			
Number Of Visitors	2	0.25%			
Repeating	2	0.25%			
Sum Of Factors Counts	789	100.00%			
# Of Factors Missed by respondents (Filtration)	72				
Total	861	Max (3 factors)	1500	Min (1 factors)	500
Number Of Factors After Merging	21	# Of Sites	21		
# Of Participants in Questionnaire	20				

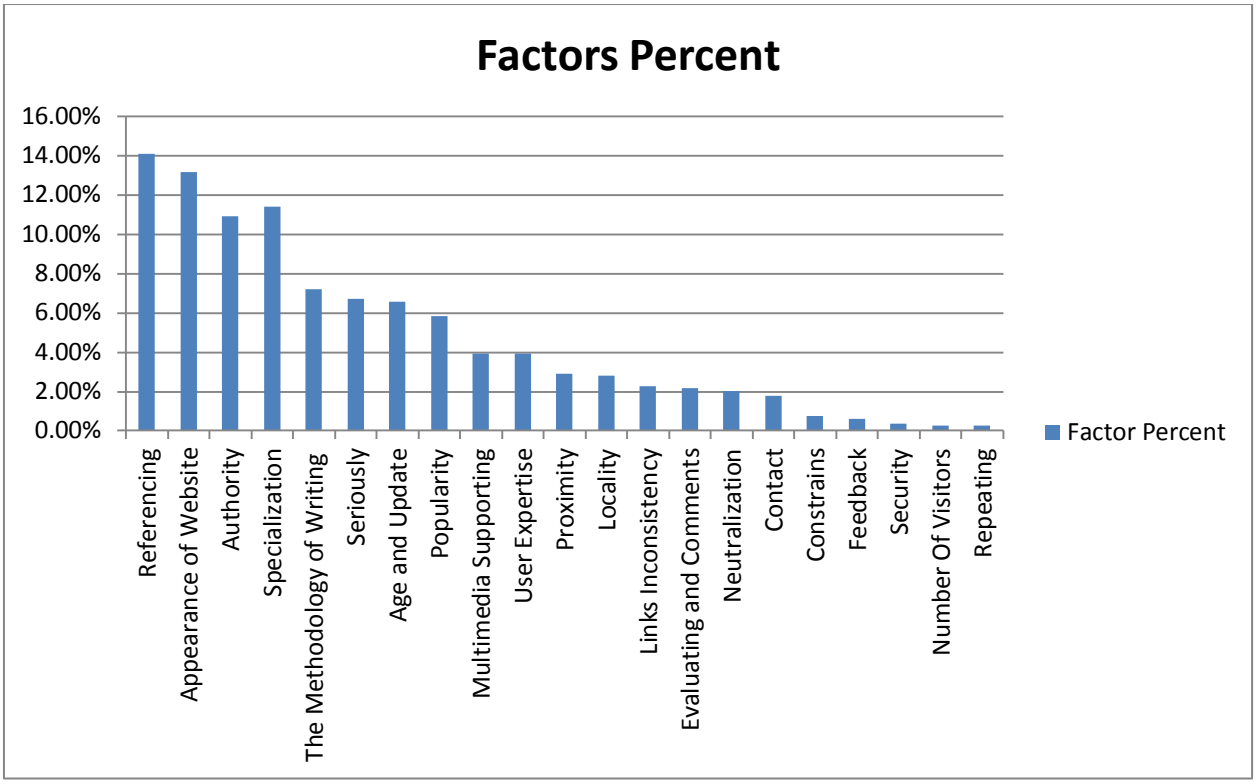


FIGURE 3. 3:THE PERCENT OF FACTORS AS RESULTED FROM THE QUESTIONNAIRE.

Figure 3.4 shows the percentage of content trust in our experiment according to category. In general we can notice how the governmental web articles come in the first place with 72% of trust from respondents' point of view. We can also notice that the existence of many strong factors in this type of sites leads to this high percentage of trust from respondents like: authority, seriousness, referencing … etc. In second and third place appear medical and scientific web articles, with a trust rate of 67% and 64.4% respectively. Specialty of these medical and scientific web pages is one of the most effecting factors on trust of their contents. News web articles come in fourth place, which is attributed to some factors that decrease the trust of content on these websites like: bias, bad or un-trusted referencing, …etc. while some factors play a role of positive support on certain types of websites like: popularity, authority, …etc. The social websites are ranked the last since they include no specialty and the referencing procedure is pure in these websites. Needless to say, one of the factors that could affect most positively on these website is the "appearance" factor.

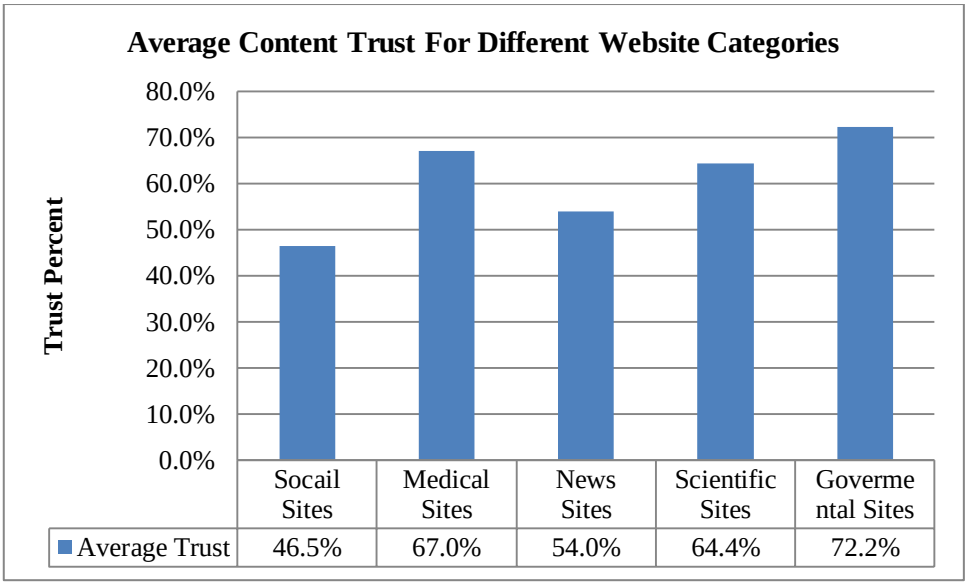


FIGURE 3. 4:CATEGORIZE RESOURCES CONTENT TRUST PERCENT.

Table 3.2 shows information resources content associated with their trust rate. Each information source content trust is calculated as the percentage of "Yes" answers to overall answers, and then we take the average content trust percentage to all information sources that belong to a specific category. The results confirm what we expected since implicit factors exist in each information resource.

TABLE 3. 2:RESOURCES CONTENT TRUST PERCENT FROM EXPERIMENT.

Site URL (Information Inside)	Yes	No	Trust %		
http://hffar.com/	8	12	40%	46.5%	Social Sites
http://www.maktoobblog.com	9	11	45%		
http://tadwen.com/archive/	12	8	60%		
http://www.khabarieh.com	9	10	47%		
http://qulqal.com/index.php?page=latest	8	12	40%		
http://www.123esaaf.com/	15	5	75%	67.0%	Medical Sites
http://www.tbeeb.net/	15	5	75%		
http://sehha.com/	13	7	65%		
http://khayma.com/tagthia/	12	8	60%		
http://www.arabmedmag.com/	12	8	60%		
https://news.google.com/news?ned=ar_me	10	10	50%	54%	News Sites
http://www.aljazeera.net/portal	14	6	70%		
http://www.almanar.com.lb/main.php	14	6	70%		
http://www.chihab.net/	9	11	45%		
http://www.alarabnews.com/	7	13	35%		
http://scitech-ar.blogspot.com/	18	2	90%	64%	Scientific Sites
http://www.arab-ency.com/index.php?t=1	17	3	85%		
http://www.embalmart.com/	8	11	42%		
http://5raeb.blogspot.com/	8	12	40%		
http://olom.info/mgz/	13	7	65%		
http://gaca.palestineit.com/Default.aspx	15	6	71%	72%	Governmental Sites
http://www.hebron-city.ps/	12	8	60%		
http://www.minfo.ps/arabic/index.php?pagess=home	15	5	75%		
http://www.moe.gov.ps/	18	1	95%		
http://www.palestinecabinet.gov.ps/Default.aspx	12	8	60%		

For instance the article in "*news.google.com*" at first glance the reader of its trust percentage of 50% will view it as low, depending on factors like popularity, authority, …etc. But in fact this article has equal factors against this opinion like bias, appearance … etc. Another thing we can notice from this result is the role of category (specialization) of resource.

3.4.2 Factors Prioritization

We want to find a mechanism to prioritize these factors empirically. First, we will study how categorization of resources affects factors order. In chapter 5, we will develop an order for the factors depending on weight formula that driven to be used as a feature for factors classification.

The order of factors will change according to each category, the table 3.3 (a, b, c, d, e) shows how these factors ranked for all five categories in our experiment:

- *Social website contents*

The order of factors in table 3.3 shows that most of respondents agreed that appearance of a website factor is coming first for social sites, the appearance factor represents to them a good indicator that makes them feel more trusting towards information inside social resources. Specialization comes in second place after appearance, which means that these resources not customized in their information that makes any of their content un-trusted. Referencing also is poor in social resources and takes third place in their attention, while authority represents to them a good indicator to trust the content if referencing is weak. The other factors have low mention from respondents that in general have low effect on social websites.

TABLE 3. 3:FACTORS ORDER FOR SOCIAL SITES CATEGORY.

	Factors	Count	Percent
Social Sites	Appearance of Website	20	23.81%
	Specialization	11	13.10%
	Referencing	10	11.90%
	Authority	10	11.90%
	The Methodology of Writing	5	5.95%
	Seriously	5	5.95%
	User Expertise	4	4.76%
	Links Inconsistency	4	4.76%
	Age and Update	3	3.57%
	Popularity	3	3.57%
	Feedback	3	3.57%
	Multimedia Supporting	2	2.38%
	Proximity	1	1.19%
	Evaluating and Comments	1	1.19%
	Neutralization	1	1.19%
	Constrains	1	1.19%
	Total	**84**	**100.00%**

- *Medical website contents*

Here we can note that specialization comes in first place for medical resources that indicate that specialty of resource means specialty of its content, which leads to more trust. Appearance and multimedia supporting are necessary for medical resource to increase the trust of their content; authority of resource becomes important as third reason generating content trust of medical resources. Users expertise factor has its effect on medical information either the respondents have prior knowledge about the information or they haven't. The other factors appears in medical websites as shown in figure 3.4

TABLE 3. 4:FACTORS ORDER FOR MEDICAL SITES CATEGORY.

	Factors	Count	Percent
Medical Sites	Specialization	20	22.73%
	Appearance of Website	16	18.18%
	Authority	9	10.23%
	User Expertise	9	10.23%
	Referencing	7	7.95%
	The Methodology of Writing	6	6.82%
	Multimedia Supporting	6	6.82%
	Popularity	4	4.55%
	Seriously	3	3.41%
	Evaluating and Comments	3	3.41%
	Age and Update	2	2.27%
	Proximity	2	2.27%
	Constrains	1	1.14%
	Total	88	100.00%

- *News website contents*

Referencing is the factor that most respondents agreed to be the most important factor affecting news websites type since it represents to them a strong indicator to trust information. Authority is another strong factor from the respondents point of view, which comes in second place leads to high trust value in their contents. Appearance of news websites that have a professional look provides to them more trust to its articles and information besides its specialty on world, sport, …etc. News, where both two factors come on third and fourth places respectively. Popularity of the news websites have certain effects on their contents trust as is the case with the neutralization factor that every news website must possess as shown in table 3.5.

TABLE 3. 5:FACTORS ORDER FOR NEWS SITES CATEGORY.

	Factors	Count	Percent
News Sites	Referencing	13	15.12%
	Authority	12	13.95%
	Appearance of Website	11	12.79%
	Specialization	8	9.30%
	Popularity	7	8.14%
	Neutralization	7	8.14%
	Seriously	6	6.98%
	User Expertise	4	4.65%
	The Methodology of Writing	4	4.65%
	Multimedia Supporting	4	4.65%
	Proximity	3	3.49%
	Locality	3	3.49%
	Age and Update	2	2.33%
	Links Inconsistency	2	2.33%
	Total	**86**	**100.00%**

- *Scientific website contents*

Most respondents express their desire to see scientific websites without science fiction information; this is done by providing a logical or scientific proof for everything written inside. Specialization, referencing, appearance and authority of scientific websites are viewed as critical factors that effect on their content as shown in table 3.6.

TABLE 3. 6:FACTORS ORDER FOR SCIENTIFIC SITES CATEGORY.

	Factors	Count	Percent
Scientific Sites	Seriously	15	17.65%
	Specialization	12	14.12%
	Referencing	10	11.76%
	Appearance of Website	10	11.76%
	Authority	8	9.41%
	The Methodology of Writing	6	7.06%
	Popularity	5	5.88%
	Multimedia Supporting	5	5.88%
	User Expertise	4	4.71%
	Age and Update	4	4.71%
	Locality	2	2.35%
	Links Inconsistency	2	2.35%
	Contact	2	2.35%
	Total	**85**	**100.00%**

- *Governmental website contents*

The All respondents agreed that governmental websites content are authorized by its nature, which provides some trust feeling to the respondents. This feeling of trust in content is increased since governmental websites characterized by seriousness, too.

TABLE 3. 7:FACTORS ORDER FOR GOVERNMENTAL SITES CATEGORY.

	Factors	Count	Percent
Governmental Sites	Authority	23	27.06%
	Seriously	11	12.94%
	Referencing	10	11.76%
	Appearance of Website	8	9.41%
	Popularity	7	8.24%
	Specialization	5	5.88%
	User Expertise	5	5.88%
	Locality	4	4.71%
	Age and Update	3	3.53%
	Neutralization	3	3.53%
	Multimedia Supporting	2	2.35%
	Proximity	2	2.35%
	The Methodology of Writing	1	1.18%
	Links Inconsistency	1	1.18%
	Total	**85**	**100.00%**

Referencing in content is requested to make the respondents more confident about some information in these websites. Appearance of governmental websites that have official look (i.e. official logo, colors, pictures…etc.) and popularity also increased trust in content of these websites. Finally when these governmental websites are dedicated in some field like: politics, sport, education or interior ministries make them more trustworthy than general governmental ones. It could be hard to trace each category of web resources over Internet and determine the order of their factors for each one empirically. Building a model depends on all categories and needs more time and a lot of work. In our work we will trace the five categories studied empirically. The final comparison for each factor according to website categories is shown in figure 3.5 that shows the percent of factors mentioned for each website category.

The question now is how we can design a model to trust in resources content based on what we analyzed? We can answer this question from the data we have from empirical experiment through two axes:

1) We can derive from the results of the experiment the number of respondents choosing this factor affecting content trust from their point of view and we will called this *factor counts*.
2) The respondents that answer the questionnaire were asked to rank the most important three factors from their opinion. Many of respondents ranked all their answers not only the weightiest three. We can use this ranking to design trust model depending on *factor weights (we will explain it deeply in chapter 5)*.

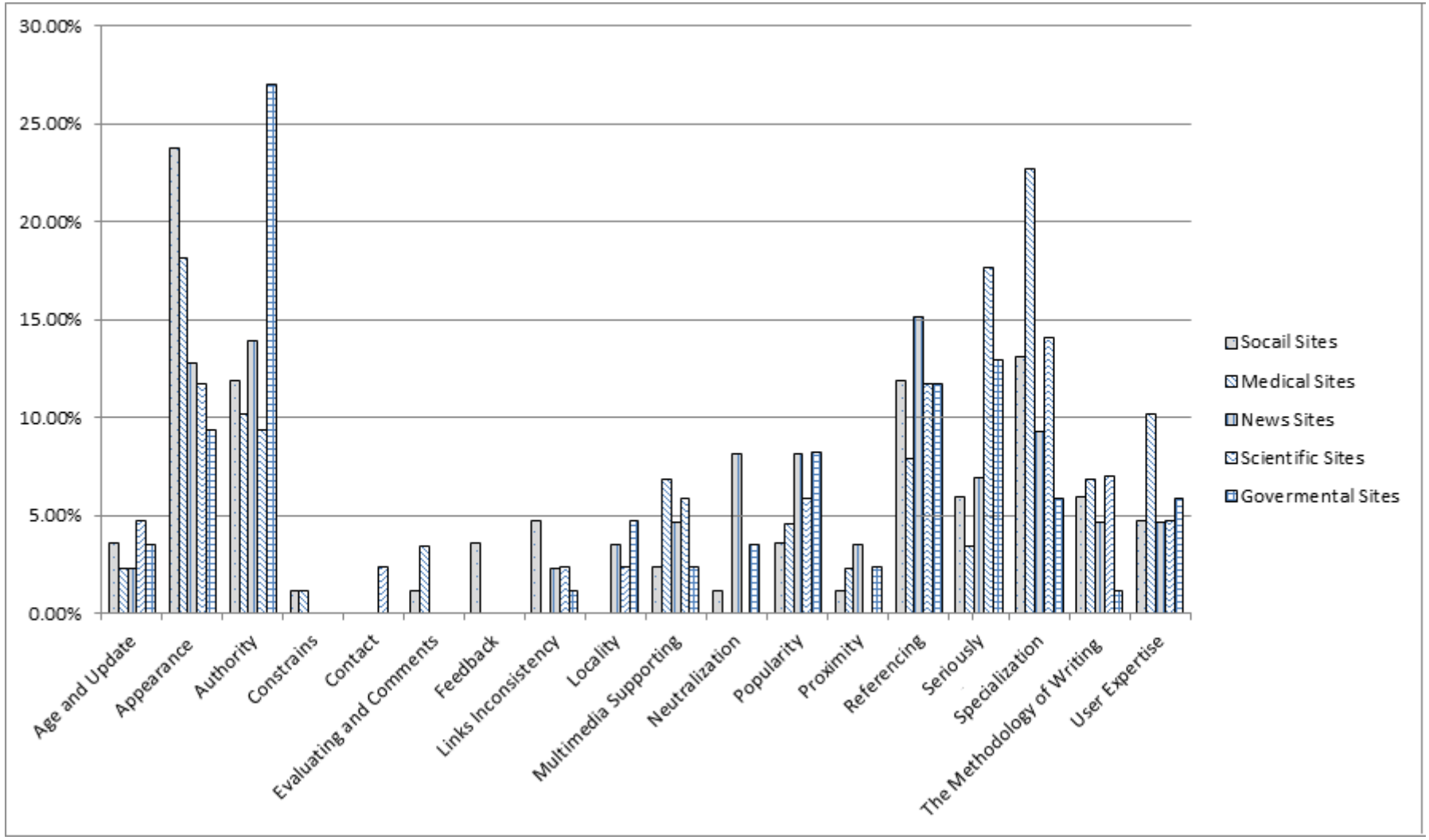


FIGURE 3. 5:PERCENT OF FACTORS MENTIONED FOR EACH WEBSITE CATEGORY.

Chapter Four

Factor Analysis Agents

Chapter 4: Factor Analysis Agents

4.1 Introduction

In this chapter, we will define *factor analysis agents, Content Trust Agent (CTA agent)* and focus on how to prepare and design *factors analysis agents* needed to evaluate factors values v_f and communicate with *CTA agent*.

4.2 Factor Analysis and CTA Agents Environment

Each factor identified in our empirical experiment must be represented with a *value* (v_f) that is captured from the *factors analysis agent*. The factor analysis agents we designed could be depending on databases, services, and automatic capture software or human being information to get and update v_f. Another type of agents are those we have called *Content Trust Agent (CTA)* dealing with trust model *(we will look in depth about CTA agent model and algorithm in chapter 5 and chapter 6)*. The figure 4.1 shows how factors analysis agents collaborate with CTA agent. Here there may some interdependence among these factor agents. However we will discuss two cases of dependences in evaluation *(chapter 7)* to improve our MAS collaboration and results. We make an assumption that our factor analysis agents are relatively independent on their work among each other while the relation between these agents need deep study to improve the model in future work.

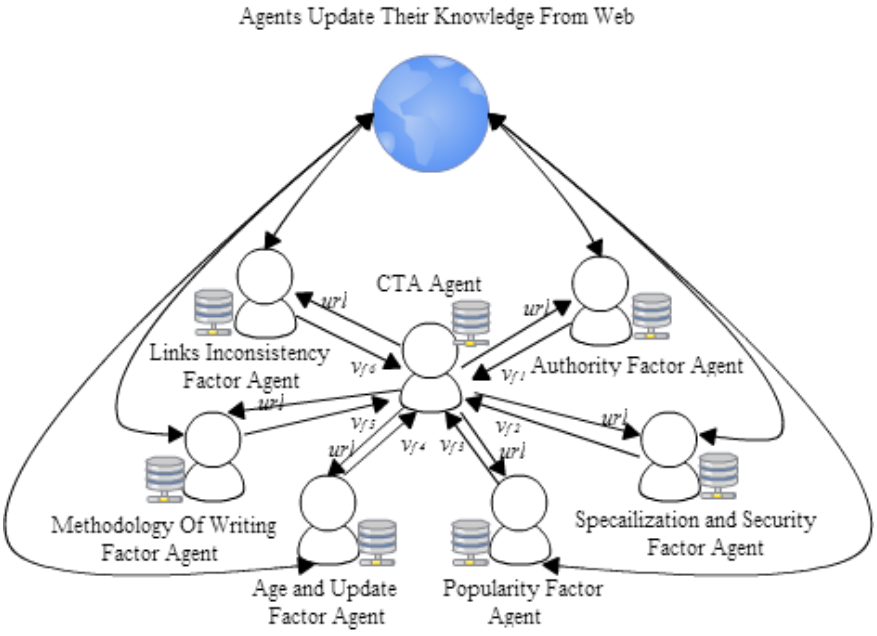


FIGURE4.1:CTA AND FACTOR AGENTS ENVIRONMENT.

CTA agent communicates with other factor analysis agents by message passing *(the detailed communication methods discussed in chapter 6)*. An agent that wants to communicate with another agent first has to create a message object, and then send it to the target agent. A message object has a kind and an optional argument object. The receiver agent determines what to do by checking the kind of received message and getting parameters as the argument object in the handleMessage() method. The information source *URL* is sent from CTA agent to all factors analysis agents asking for their values v_f about request URL. They returned the values about requested URL based on their knowledge and its calculation to be used in our trust model inside CTA agent. We will explain in this chapter how we design and prepare each factor analysis agent to collaborate with our CTA agent and how we implement them. In chapter 6 we will study the interactions of these autonomous agents with a view to assessing their effects on the system model as a whole.

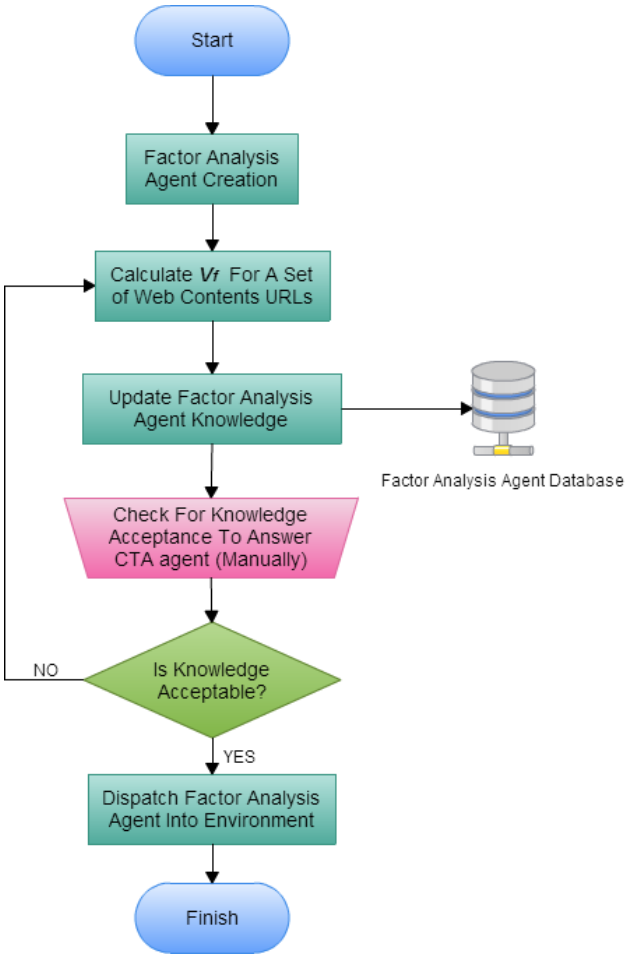


FIGURE4.2:FLOWCHART FOR PREPARATION OF FACTOR ANALYSIS AGENT TO BE DISPATCHED TO AGENTS ENVIRONMENT.

Each factor analysis agent has its autonomous work to extract and accumulate their own database (knowledge) continuously, but before that we checked manually if the knowledge of each factor

analysis agent becomes acceptable to answer CTA agents after we created them. They could take a very long calculation time to achieve that sometimes hours, days even months. After they have acceptable knowledge we dispatch them into our environment that appears in figure 4.1 to be ready to answer CTA agent and update their knowledge automatically. The whole process of preparing any factor analysis agent to be dispatched is represented in flowchart in figure 4.2.

After dispatching factor analysis agents to agents' environment they start doing two things:

- *Answering CTA agent requests:* Each time CTA agent sent a request message about a specific webpage content URL, all factor analysis agents are always ready to respond to this request based on their updated knowledge while direct calculation for every request takes very long calculation time could reach to months as we mentioned before.
- *Keep updating their knowledge:* The factor analysis agents keep continuously and sufficiently updating their knowledge. The process of updating still needs a periodic feedback to check knowledge acceptance.

4.3 Factor Analysis Agents Implementation

4.3.1 Authority Factor Agent

There are many algorithms that provide a definition for authority; one of the most popular algorithms is PageRank. A PageRank results from a mathematical algorithm based on the web graph, created by all World Wide Web pages as nodes and hyperlinks as edges, taking into consideration authority hubs such as cnn.com or usa.gov. [27]

In our work we used PageRank service in order get authority factor value, it provides ranking values from 0 to 10 for every URL query to be used in our next evaluation experiment.

4.3.2 Specialization and Security Factors Agent

There is no reliable service that depends on specialty term for websites; on other hand there exist many standard services that categorized websites. Categorization is the process in which ideas and objects are recognized, differentiated, and understood. Categorization implies that objects are grouped into categories, usually for some specific purpose [26]. So, our work will use Categorization term to be equivalent to Specialization.

In this research, we used standard values to evaluate categorization and security. We used McAfee TrustedSource™ database [3] for our work, we got permission through Email from the company to use their database with maximum number of 100 URLs per file at a time (*See Appendix B*).

The TrustedSource Web Database uses categories to organize similar types of URLs into groups based on the content of the web page. For example, www.mcafee.com, www.trustedsource.org, and www.webwasher.com are grouped into the Business category. We used *SmartFilter XL* Database that is comprised of 91 categories. [22]

Security factor called by TrustedSource web database as *Reputation* that uses an automated process that looks at many different security attributes of that URL - the URL's content, where the URL is showing up on the Internet, the URL's domain behavior, and more. TrustedSource determines a score that represents the risk to your network, computers, and personal information when you visit any URL. The reputation score is represented in the database regardless of its categorization status. This means a URL can have a web reputation score, but is not in any category [22].

We use our reputation score for each different type of risk as shown in table 4.1:

TABLE 4.1:REPUTATION SCORES.

Risk type	Web reputation score
Minimal Risk	9
Unverified	7
Medium	5
High	1

4.3.3 Popularity and Number Of Visitors Factors Agent

The popularity of web sites and individual pages or sections within a site is measured by web traffic, Web traffic is the amount of data sent and received by visitors to a web site. Number of visitors is one type of information that often collated when monitoring web traffic, so we merge this factor with popularity in our model [29].

In our work we used Alexa database itself for popularity factor, since it have 1,000,000 top sites that are ordered by popularity [7].

4.3.4 Age and Update Factor Agent

As we mentioned in chapter2 *age and update* are frequent updating of content, at least once every few weeks, and more often, once a week or more. There is no helpful service or database that monitors webpage content update or changes, so we develop our agent to automatically capture this factor depends on monitor changes for any webpage URL. The algorithm of age and update factor depends on two inputs: webpage URL that we want to check its content update and standard update time. The following represents the algorithm steps:

ALGORITHM 4.1 .:ONLINE CAPTURE AGE AND UPDATE FACTOR

Require: webpage URL: *wp_url*, standard updating time: *t*
1. Create connection to *wp_url.*
2. **for** *i*=1 To 10 **do**
3. Last_Str ← *SHA-256 digest* (*wp_url*)
4. Wait (t/10)
5. Current_Str ← *SHA-256 digest* (*wp_url*)
6. **if** Last_Str *Not equals* Current_Str **then**
7. Increase update *rank counter*
8. **end if**
9. **end for**
10. Store *rank counter* in age and update database

SHA-256 digest (wp_url) function depends on a family of cryptographic hash functions published by the National Institute of Standards and Technology (NIST) called Secure Hash Algorithm (SHA). SHA-2 a family of two similar hash functions, with different block sizes, known as SHA-256 and SHA-512. They differ in the word size; SHA-256 uses 32-bit words where SHA-512 uses 64-bit words [2, 28].

The algorithm simply checks and monitors any changes in hashed webpage content through using SHA-*256 digest (wp_url) function* result before and after passing *t/10*, where *t* is standard update time. The time *t/10* because we want the final result of the algorithm a rank between 0 to 10 according to overall standard updating time. Finally the agent store its information about age and update in its own database and used them if necessary.

4.3.5 Methodology Of Writing Factor

This factor can be implemented by using spelling and grammar checking for a few or no errors in webpage content. We make an assumption that no page has more than three misspelled words or four grammatical errors. Here we wish to acknowledge that this tool and service for this factor are customized for webpages written in English language.

- **Grammar checking**

In order to check for grammatical errors, we need a tool that helps to achieve that. "LanguageTool" is a tool that comes with its own embedded HTTP server so we can send a text to "LanguageTool" via HTTP and get the detected errors back as XML. This embedded server can be started using stand-alone application *(see figure 4.3)* and configure it to listen on a port that is not used yet, also be available in server mode until we stop it [20]

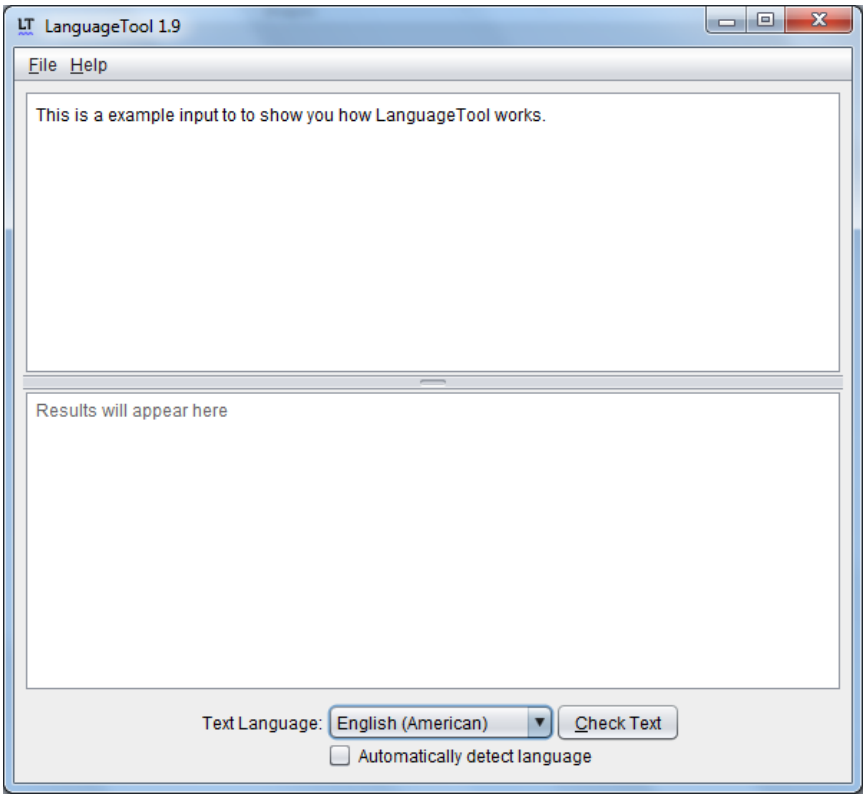


FIGURE4.3:LANGUAGETOOL APPLICATION.

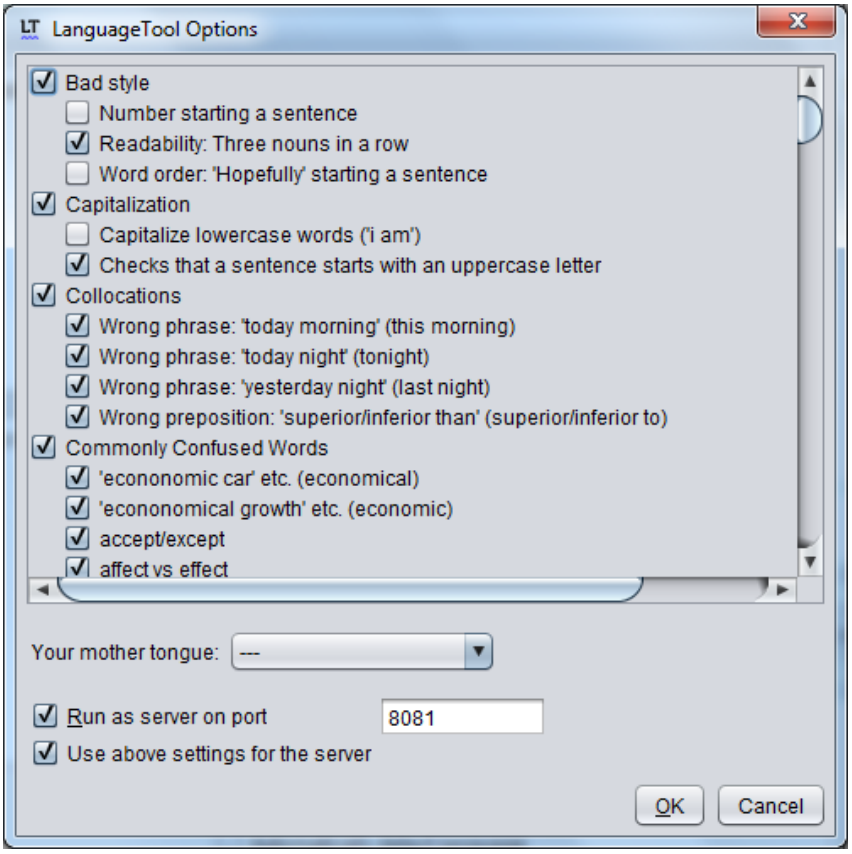


FIGURE4.4: LANGUAGETOOL OPTIONS.

The algorithm for checking grammar of webpage content as following:

ALGORITHM 4.2 .:ONLINE CHECK GRAMMAR OF WEBPAGE CONTENT

Require: webpage URL: *wp_url,*

1. Create connection to *wp_url.*

2. *C ← URLTagContent (("P","TD"),wp_url)*

3. **for** *i*=1 To *Size(C)* **do**

4. *xml_errors ← CheckErrors* (*Ci*)

5. **if** *xml_errors Not equals* 0 **then**

6. Increase Error Counter *Ec*

7. **end if**

8. **end for**

9. **if** *Ec* > 4 **then**

10. *Grammar_Rank =Round (Ec/Size(C)*10)*

11. else

12. *Grammar_Rank =10*

13. end if

14. Store *Grammar_Rank* temporarily

The algorithm started by establishing a connection to webpage content URL, the function in line 2 is used to extract texts inside <P> and <TD> tags and store these texts into C array. Grammar checker algorithm starts checking errors for each element in C, these errors are returned as XML. The element that has an error makes error counter *Ec* increases, when C elements finished we check our assumption that no webpage has more than 4 grammatical errors (line9- line13). Finally we temporarily store the results to be merge later with *Spelling_Rank* value.

- **Spell Checking:**

We used Jazzy the java open source spell checker that allows us to add spell-checking functionality to Java Applications easily. The Jazzy team plans to create an applet that would be suitable for use in HTML forms that will allow the user to spell check html text area before submitting [17].

We used Jazzy in our agent to check for any spelling errors in webpages contents, the algorithm that used to represent *Spelling_Rank* value depends on the ratio between number of misspelled words exists in a webpage to the whole number of words.

ALGORITHM 4.3 .:ONLINE CHECK SPELLING OF WEBPAGE CONTENT

Require: webpage URL: *wp_url,*
15. Create connection to *wp_url.*
16. *C ← URLTagContent (("P","TD"),wp_url)*
17. **for** *i*=1 To *Size(C)* **do**
18. *W ←* *WordTokenizer* (*Ci*)
19. **for** *j*=1 To *Size(W)* **do**
20. **if** *checkSpelling(Wj) Not True* **then**
21. Increase Misspelled Counter *MSc*
22. **end if**
23. *whole_words ← whole_words+Size(Wj)*
24. **end for**
25. **end for**
26. **if** *MSc* > 3 **then**
27. *Spelling_Rank =Round((1-(MSc/ whole_words))*10)*
28. **else**
29. *Spelling_Rank =10*
30. **end if**
31. Store *Spelling_Rank* temporarily

We finally take *Grammar_Rank* and *Spelling_Rank* values to generate final results for methodology of writing factor.

4.3.6 Links Inconsistency Agent:

We used Link Checker application to implement links inconsistency factor that accepts a website page URL as input then scanned for all the elements which points to other URLs, images, files etc. A collection of all Links are then scanned further and classified into 'good' and 'broken' links and provide *page_health* value that used by out agent [21].

4.3.7 Appearance Of Website Agent:

There are no services that help us evaluate this factor; appearance of webpage cannot be easily implemented. Moreover, it needs a sophisticated algorithm that captures many of webpage design expert opinions and transforms these opinions into programming codes. We started implementing this factor, but a lot of issues appeared during the work. For instance, based on expert advice and results in work [8] "*There is evidence that children prefer sans serif fonts (Arial & Comic) over serif fonts*. " This result, however, cannot be generalized for all websites since there exist many children who speak other languages.

We make an assumption for algorithm for English websites only, relying on design expert advices. The model we proposed starts reading HTML for webpage to merge with the three types of CSS: inline, internal and external order based on their priorities (from high to low), the result of merging store in XML file contains all necessarily design tags to transfer expert advice to program (*called Style Store XML SSXML*), figure 4.6 shows the architecture of XML file content.

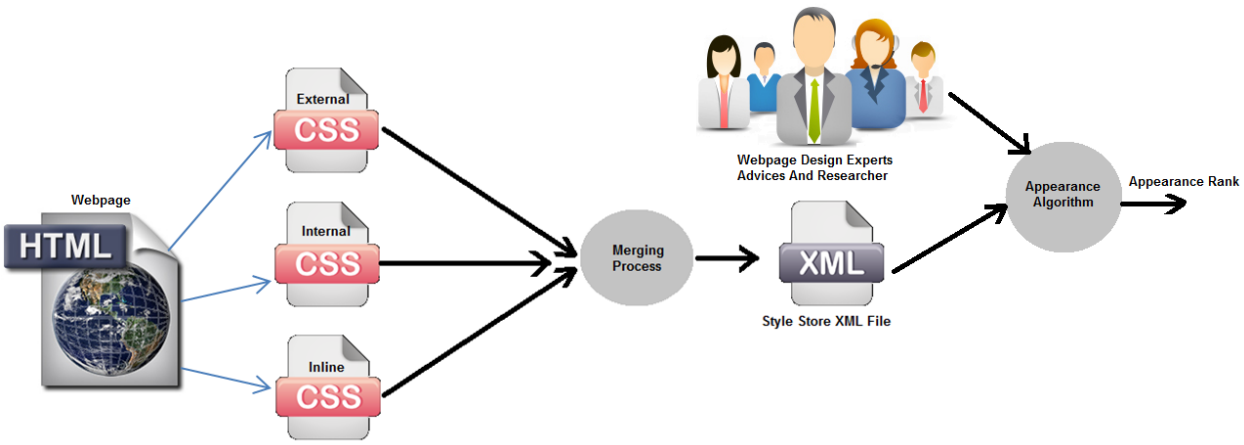


FIGURE4.5:CSS MERGING PROCESS.

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<stylestore>
    <style>
      <tagname>BODY</tagname>
      <tagid>*</tagid>
      <tagclass>*</tagclass>
      <tagcolor>000000</tagcolor>
      <tagbgcolor>null</tagbgcolor>
      ...
    </style>
    ...
</stylestore>
```

FIGURE4.6 .:ARCHITECTURE OF STYLE STORE XML FILE.

After generating SSXML we are ready to transform any design expert advice, when webpage design opinions increase the appearance algorithm will provide more precise results *(see figure 4.5)*.

The drawbacks of this technique are: first it needs a lot of time to program one opinion. Second, it takes a very long execution time especially when expert opinions increased. So we build an autonomous agent specialized in webpage appearance value extraction *(have its own database)* that can communicate with our CTA agent.

Chapter Five

Content Trust Agent (CTA) Model

Chapter 5: Content Trust Agent (CTA) Model

5.1 Introduction

In this chapter, we propose a new website content trust model that represents the heart of CTA agent to generate a trust value based on factor analysis agents. We provide a mathematical formula and a proof requirement to satisfy the model algorithm properties based on classification of factors' of content trust. In addition to that, we describe some examples about the trust model algorithm.

5.2 Proposed Model

The model we proposed depends on classifying respondents' answers of empirical experiment. We need a feature to rely on and show importance of these identified factors and prioritize them. We used two features to make the classification: *Factors weight* and *factors count*. The factor count represents by summing the number of respondents choosing this factor that affects content trust from their point of view *(you can see table 3.1 in chapter 3)*. Factors weights can be derived from respondents' answers when they ranked most important three factors (some of them ranked over 3 factors). This ranking of factors for each respondent indicates how important these factors to them. We design in the next section a mathematical formula to maintain the weight of factors from respondents' point of view.

5.3 Factors Weights Mathematical Formula

Let f indicate a factor, $(f_1, f_2, f_3, \ldots f_n)$ are a set of ***n*** factors that exist in our trust model M, w_f is factor weight. f_i^j represents number of times a factor ***i*** ranked as ***j***, $j \in \{1, 2, \ldots K\}$, $i \in \{1, 2, \ldots n\}$ where K is the upper limit of ranking. R^j is a partial weight for all factors ranked as ***j***. So, for any factor *i*

$$R^j = \frac{\sum_{i=1}^{n} f_i^j}{\sum_{m=1}^{K} \sum_{i=1}^{n} f_i^m}$$

$$\text{for each } j \in \{1, 2, \ldots K\}$$

And its weight is:

$$w_{f_i} = f_i^1 R^1 + f_i^2 R^2 + \cdots + f_i^K R^K$$

The above formula can be abbreviated into the following formulas:

$$w_{f_i} = \sum_{j=1}^{K} f_i^j R^j$$

where $\sum_{j=1}^{K} R^j = 1$, for each factor $n \in \{1,2 \dots N\}$

The final formula of w_{f_i} has the ability to give a feature of importance in addition to *factor count* feature. In this stage we can find the factors' weights that we have from empirical experiment. Table 5.1 shows the results of experiment to find f_i^j for each factor ***i***.

TABLE 5.1:WEIGHTS FORMULA RESULTS FOR EXPERIMENT FACTORS.

i	Factor (*f*)	Count	Factor Per.%	f_i^1	f_i^2	f_i^3	f_i^4	f_i^5	f_i^6		
1	Referencing	111	14.07%	52	41	13	4	1	0	w_{f_1}	**41.789608**
2	Appearance of Website	104	13.18%	63	25	10	6	0	0	w_{f_2}	**42.850113**
3	Authority	91	11.53%	62	19	9	1	0	0	w_{f_3}	**40.250213**
4	Specialization	90	11.41%	59	25	4	2	0	0	w_{f_4}	**39.795313**
5	The Methodology of Writing	57	7.22%	22	28	7	0	0	0	w_{f_5}	**20.869813**
6	Seriously	53	6.72%	40	8	3	2	0	0	w_{f_6}	**24.454413**
7	Age and Update	52	6.59%	14	18	14	6	0	0	w_{f_7}	**14.726213**
8	Popularity	41	5.20%	26	11	3	1	0	0	w_{f_8}	**17.685813**
9	Multimedia Supporting	31	3.93%	15	7	6	2	1	0	w_{f_9}	**10.970413**
10	User Expertise	31	3.93%	25	3	1	2	0	0	$w_{f_{10}}$	**14.631113**
11	Proximity	23	2.92%	9	7	4	0	2	1	$w_{f_{11}}$	**7.399713**
12	Locality	22	2.79%	10	8	3	1	0	0	$w_{f_{12}}$	**8.138813**
13	Links Inconsistency	18	2.28%	9	4	4	1	0	0	$w_{f_{13}}$	**6.560713**
14	Evaluating and Comments	17	2.15%	4	5	6	1	1	0	$w_{f_{14}}$	**4.384313**
15	Neutralization	16	2.03%	11	4	1	0	0	0	$w_{f_{15}}$	**7.243913**
16	Contact	14	1.77%	2	6	4	2	0	0	$w_{f_{16}}$	**3.381813**
17	Constrains	6	0.76%	2	1	2	1	0	0	$w_{f_{17}}$	**1.655413**
18	Feedback	5	0.63%	3	1	1	0	0	0	$w_{f_{18}}$	**2.036913**
19	Security	3	0.38%	0	3	0	0	0	0	$w_{f_{19}}$	**0.867013**
20	Number Of Visitors	2	0.25%	0	2	0	0	0	0	$w_{f_{20}}$	**0.578013**
21	Repeating	2	0.25%	0	2	0	0	0	0	$w_{f_{21}}$	**0.578013**
	Sum Of Factors Counts	**789**	**100.00%**	**428**	**228**	**95**	**32**	**5**	**1**		**789**

- $\boldsymbol{f_i^1}$: The summing of factors counts that appear as number 1,which is equal 428 in table 5.1.
- $\boldsymbol{f_i^2}$: The summing of factors counts that appear as number 2,which is equal 228 in table 5.1.
- The same for $\boldsymbol{f_i^3}$, $\boldsymbol{f_i^4}$, $\boldsymbol{f_i^5}$ **and** $\boldsymbol{f_i^6}$ that their values are 95, 32, 5 and 1 respectively.
- $\boldsymbol{R^1}$: Is a partial weight for $\boldsymbol{f_i^1}$ calculated as $R^j = \frac{\sum_{i=1}^{n} f_i^j}{\sum_{m=1}^{K}\sum_{i=1}^{n} f_i^m} = \frac{428}{789} = 0.54245881$, where $\boldsymbol{R^2} = \frac{\boldsymbol{228}}{\boldsymbol{789}} = 0.2889734$, $\boldsymbol{R^3} = \frac{\boldsymbol{95}}{\boldsymbol{789}} = 0.12040558$, $\boldsymbol{R^4} = \frac{\boldsymbol{32}}{\boldsymbol{789}} = 0.04055767$, $\boldsymbol{R^5} = \frac{\boldsymbol{5}}{\boldsymbol{789}} = 0.00633714$, $\boldsymbol{R^6} = \frac{\boldsymbol{1}}{\boldsymbol{789}} = 0.00126743$.
- To calculate the weight of factor 1 we used $\boldsymbol{w_{f_1}} = \sum_{j=1}^{K} f_1^j R^j = (52 \times 0.54245881) + (41 \times 0.2889734) + (13 \times 0.12040558) + (4 \times 0.04055767) + (1 \times 0.00633714) + (0 \times 0.00126743) = 41.78960788$

Ordering the results provides a clear vision of which of these factors deserve to be on top of the list. Besides, they have major influence on the system. The weights preserve factors' importance and prioritize, e.g. "Appearance of Website (f_2)" moves from second place to the first, because 63 persons in the experiment priorities this factor as first when they asked to prioritize the three most influencing factors as seen in figure 5.1 and table 5.2.

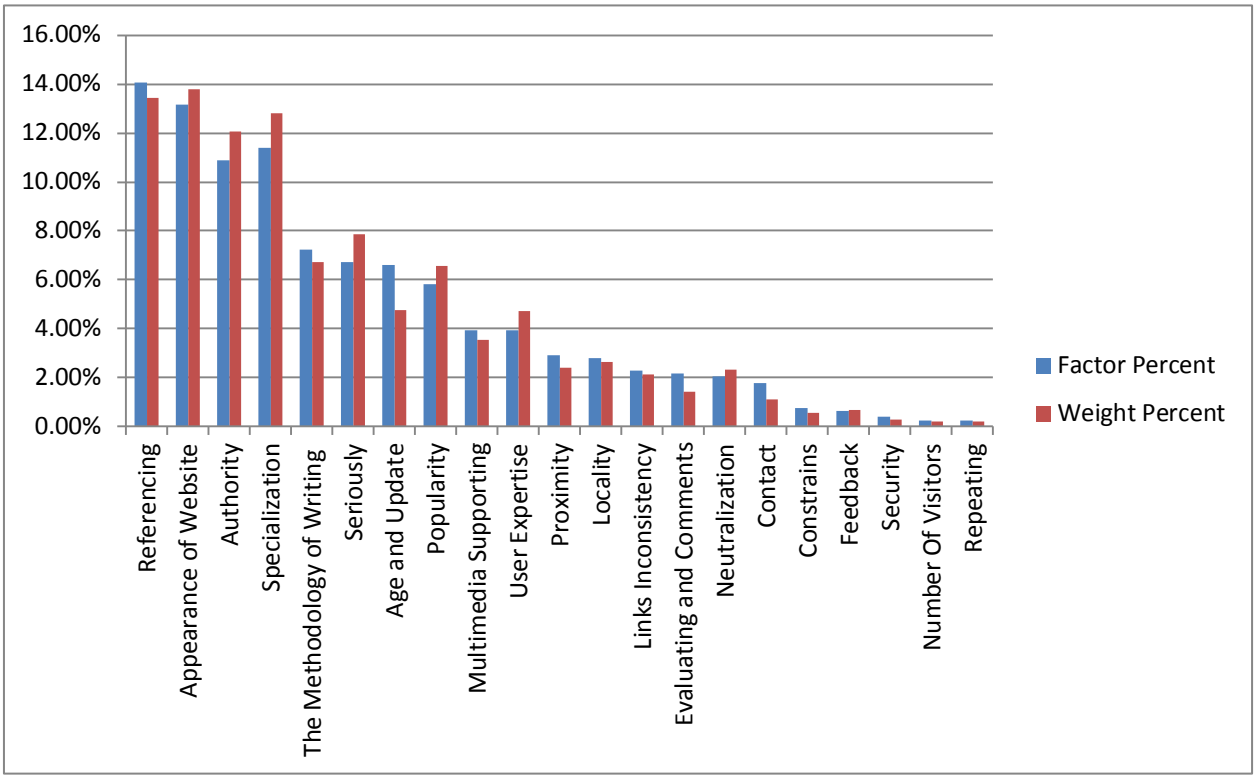


FIGURE 5.1:FACTORS WEIGHTS VS. WEIGHT PERCENT FEATURES

5.4 Factors Classification

After we determine the two features, we are ready to use them in our classification process, figure 5.2 shows how the classification results appears and how we classify the factors into three bands:

- High Rational (HR):

This band of factors is classified to be the most important ones includes *(Referencing, Appearance, Authority and Specialization)*. This importance means that each factor has a critical effect on the final result of our model. For instance, if one of them is high in its value then the final result of our model will be high regardless of other factors values.

- Medium Rational (MR):

MR factors aim to support HR factors values besides they have a medium effect in final model result. These supportive factors are: *The Methodology of Writing, Seriously, "Age and Update" and Popularity.*

- Low Rational (LR):

This variable also plays a role of supporting factors like MR except they have a low effect on content trust model. The rest of factors are belonging to this band of factors. *See table 5.2.*

TABLE 5.2:FACTORS WEIGHTS ORDER AND CLASSIFICATION.

Factor (f)	w_{f_n}	Classification
Appearance of Website ($f2$)	42.850113	High Rational (HR)
Referencing ($f1$)	41.792913	
Authority ($f3$)	40.250213	
Specialization ($f4$)	39.795313	
Seriously ($f6$)	24.454413	Medium Rational (MR)
The Methodology of Writing ($f5$)	20.869813	
Popularity ($f8$)	17.685813	
Age and Update ($f7$)	14.726213	
User Expertise ($f10$)	14.631113	Low Rational (LR)
Multimedia Supporting ($f9$)	10.970413	
Locality ($f12$)	8.138813	
Proximity ($f11$)	7.399713	
Neutralization ($f15$)	7.243913	
Links Inconsistency ($f13$)	6.560713	
Evaluating and Comments ($f14$)	4.384313	
Contact ($f16$)	3.381813	
Feedback ($f18$)	2.036913	
Constrains ($f17$)	1.655413	
Security ($f19$)	0.867013	
Number Of Visitors ($f20$)	0.578013	
Repeating ($f21$)	0.578013	

Why did we use the term *"Rational"* in our classification? We used this term because the factors are identified from respondents' answers during empirical experiment without any biasing to these factors, so the respondents answered based on their rational senses.

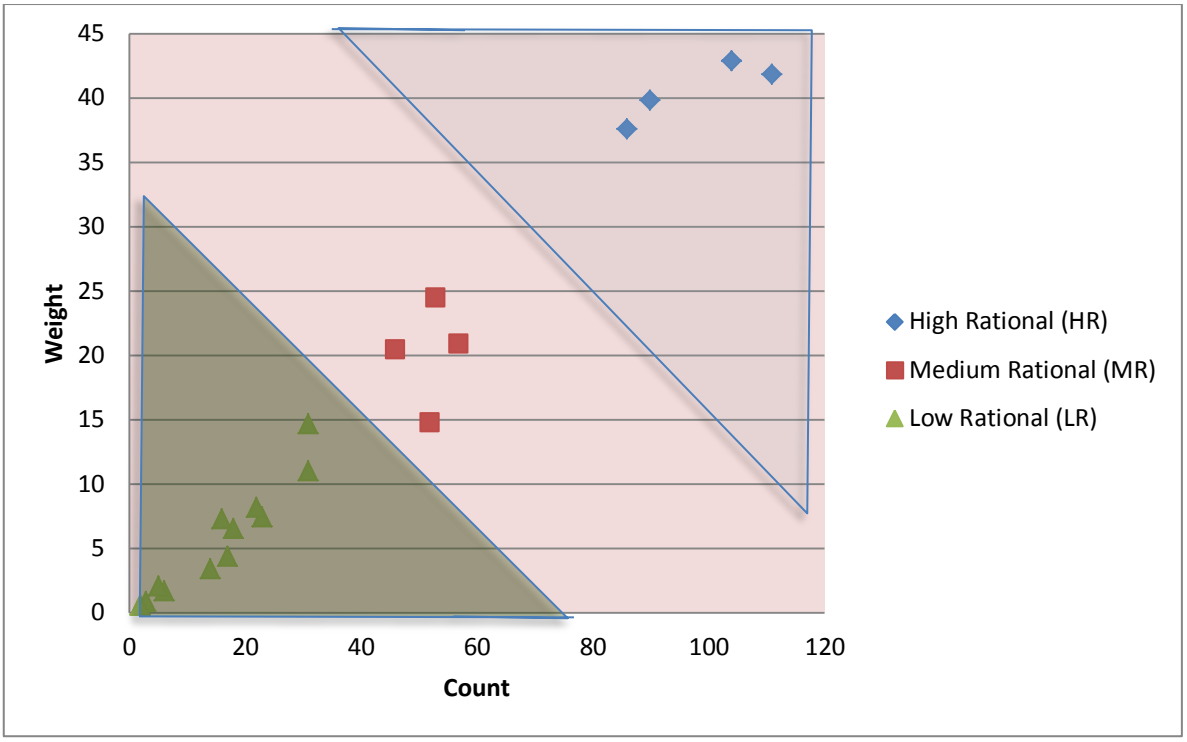


FIGURE 5.2:FACTORS CLASSIFICATION ACCORDING TO FACTORS COUNT AND WEIGHT FEATURES.

5.5 Content Trust Model Algorithm

The model *M* we designed gives trust value as a percent between 0% and 100% (*0% is un-trusted content at all, 100% completely trusted*), now we want to design a trust model algorithm using the three bands of factors classification and preserving the importance of each band. The CTA agent sends a requesting messages containing information source URL to the factor analysis agents, with each factor analysis agent providing CTA agent with its values v_f after doing its calculation where $v_f \in \{0,1 \dots 9\}, v_{f_{min}} = 0, v_{f_{max}} = 9$ and 0 means the existence of a factor on webpage article which does not appear at all *(relatively low value)*, where 9 means that the factor completely appears in webpage article *(relatively high value)*. For example, if "Appearance" factor agent returns 0 that means the URL of webpage article does not have professional or user-friendly look at all, and if it returns 9, it means URL of webpage article completely has professional or user-friendly look. Then the model *M* applies rounding technique as a method for factors importance preservation. Each factor effect on the power of rounding process depends on what band it belongs to, i.e. HR factors effect on most significant part of content trust percentage result, MR effects on decimal part while LR effects on least decimal part of the result. *Algorithm1* shows a detailed process of content trust model.

ALGORITHM 5.1 :MODEL ALGORITHM

Require: All factors weights must belong to HR band
M: Is holding the final model result initially equal 0
$\boldsymbol{v_f}$**:** A value that represents a factor
$\boldsymbol{v_{f_i}}$**:** Factors values set, ordered descending in each band (HR, MR and LR) depending on ***i*** to determine each band borders. When:
$i \in [1, l]$ ➔ HR band, where 1 and *l* are HR borders
$i \in [l+1, m]$➔ MR band, where *l+1* and *m* are MR borders
$i \in [m+1, n]$ ➔ LR band, where *m+1* and *n* are LR borders

---------- High Rational (HR) Stage ----------

1. **for each** *factor value* $v_{f_i}, i \in [1, l]$ **do**
2. $v_{f_{max}} \leftarrow Max\ (v_{f_i})$
3. **if** v_{f_i} *Is Not Repeated* **then**
4. $M \leftarrow M + v_{f_i} \times 10^{v_{f_i}}$
5. **else**
6. $x \leftarrow availabe\ power\ \ 0 \leq x \leq (v_{f_{max}} - 1)$
7. **if** *x Is NULL* **then**
8. $M = M + v_{f_i} \times 10^{0}$
9. **else**
10. $M = M + v_{f_i} \times 10^{x}$
11. **end if**
12. **end if**
13. **end for**

---------- Medium Rational (MR) Stage ----------

14. $P \leftarrow v_{f_{max}} - 2$
15. **for each** *factor value* $v_{f_i}, i \in [l+1, m]$ *and* $P \geq 0$ **do**
16. $M \leftarrow M + v_{f_i} \times 10^{P}$
17. $P \leftarrow P - 1$
18. **end for**

---------- Low Rational (LR) Stage ----------

19. $P \leftarrow v_{f_{max}} - 5$
20. **for each** *factor value* $v_{f_i}, i \in [m+1, n]$ *and* $P \geq 0$ **do**
21. $M \leftarrow M + v_{f_i} \times 10^{P}$
22. $P \leftarrow P - 1$
23. **end for**
24. $M \leftarrow {}^{M}/_{10^{v_{f_{max}}}}$
25. End

5.6 Model Selected Examples

In this section we will take some examples in our trust model to clarify how it works. The number of combinations of 21 factors is huge, each of them could have a value from 0 to 9, and we have run our mathematics and validated using reasonable set of data is illustrated in *appendix C*.

Example #1:

Suppose factors analysis agents provide CTA agent with the following values: Note: *vf1* is the value that returns from "Referencing" factor agent, *vf2* is the value that returns from "Appearance" factor agent …etc. *see table 5.1 for the order of factor values*

vf1	*vf2*	*vf3*	*vf4*	*vf5*	*vf6*	*vf7*	*vf8*	*vf9*	*vf10*	*vf11*	*vf12*	*vf13*	*vf14*	*vf15*	*vf16*	*vf17*	*vf18*	*vf19*	*vf20*	*vf21*	*Trust Percent*
0	0	0	0	0	0	0	4	4	4	4	4	4	4	4	4	4	4	4	5	5	0.40%

The above example discusses how the model will behave if the first four values are *"very low"* that represent HR factors values too and the second four values represents MR factors values three of them *"very low"* and the four is *"medium",* while the last 13 values of LR are have all *"medium"* values.

Then the trust percentage will be:

$v_{f_{max}} = 4$, MR= $v_{f_{max}} \times 10^{v_{f_{max}}-2}$, LR=$v_f \times 10^{v_{f_{max}}-5}$, v_f Ordered Descending

MR LR

Trust Percent=($4*10^{2}+5*10^{-1}+5*10^{-2}+4*10^{-3}+4*10^{-4}+4*10^{-5}+4*10^{-6}+4*10^{-7}+4*10^{-8}+4*10^{-9}+4*10^{-10}+4*10^{-11}+4*10^{-12}+4*10^{-13}$)/$10^{4}$ = 0.00400554444444444 ≈**0.40%**

LR $10^{v_{f_{max}}}$

This result 0.40% indicates that our model generates a *very low* trust value in terms of accuracy of information.

Example #2:

Suppose factors analysis agents provides CTA agent with the following values:

vf1	*vf2*	*vf3*	*vf4*	*vf5*	*vf6*	*vf7*	*vf8*	*vf9*	*vf10*	*vf11*	*vf12*	*vf13*	*vf14*	*vf15*	*vf16*	*vf17*	*vf18*	*vf19*	*vf20*	*vf21*	*Trust Percent*
0	0	1	2	2	2	2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	21.52%

This example discusses how the model will behave if the first four values are *"low"* that represent HR factors values too and the second four values represents MR factors values three of

them *"low"* and the four is *"medium"*, while the last 13 values of LR are have all *"medium"* values.

Then the trust percentage will be:

$v_{f_{max}} = 2$, HR= $v_{f_{max}} \times 10^{v_{f_{max}}}$, MR= $v_{f_{max}} \times 10^{v_{f_{max}}-2}$, LR=$v_f \times 10^{v_{f_{max}}-5}$, v_f Ordered Descending

HR MR LR

Trust Percent=($2*10^{2}+1*10^{1}+5*10^{0}+2*10^{-1}+2*10^{-2}+2*10^{-3}+5*10^{-3}+5*10^{-4}+5*10^{-5}+5*10^{-6}+5*10^{-7}+5*10^{-8}+5*10^{-9}+5*10^{-10}+5*10^{-11}+5*10^{-12}+5*10^{-13}+5*10^{-14}+5*10^{-15}$)/$10^{2}$

LR $10^{v_{f_{max}}}$

= 0.215227555555555≈ **21.52%**

This result 21.52% indicates that our model generates a *low* trust value in terms of accuracy of information.

Example #3:

Suppose factors analysis agents provides CTA agent with the following values:

vf1	*vf2*	*vf3*	*vf4*	*vf5*	*vf6*	*vf7*	*vf8*	*vf9*	*vf10*	*vf11*	*vf12*	*vf13*	*vf14*	*vf15*	*vf16*	*vf17*	*vf18*	*vf19*	*vf20*	*vf21*	*Trust Percent*
0	0	6	3	8	8	8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	81.49%

This example discuss how the model will behave if the first two HR values are *" very low"* the third is *"medium"* and the fourth is *"low"* and the MR factors values all of them *"high"* and, while the last 13 values of LR are all have *"high"* values.

Then the trust percentage will be:

$v_{f_{max}} = 6$, HR= $v_{f_{max}} \times 10^{v_{f_{max}}}$, MR= $v_{f_{max}} \times 10^{v_{f_{max}}-2}$, LR=$v_f \times 10^{v_{f_{max}}-5}$, v_f Ordered Descending

HR MR LR

Trust Percent=($6*10^{6}+3*10^{3}+8*10^{4}+8*10^{3}+8*10^{2}+8*10^{1}+9*10^{1}+9*10^{0}+9*10^{-1}+9*10^{-2}+9*10^{-3}+8*10^{-4}+8*10^{-5}+8*10^{-6}+8*10^{-7}+8*10^{-8}+8*10^{-9}+8*10^{-10}+8*10^{-11}$)/$10^{6}$

LR $10^{v_{f_{max}}}$

= 0.609197999988≈ **60.92%**

This result 60.92% indicates that our model generates a *medium* trust value in terms of accuracy of information.

Example #4:

Suppose factors analysis agents provides CTA agent with the following values:

vf1	*vf2*	*vf3*	*vf4*	*vf5*	*vf6*	*vf7*	*vf8*	*vf9*	*vf10*	*vf11*	*vf12*	*vf13*	*vf14*	*vf15*	*vf16*	*vf17*	*vf18*	*vf19*	*vf20*	*vf21*	*Trust Percent*
0	6	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	89.59%

This example discusses how the model will behave if the first HR value is *" very low"*. The second is *"medium"* and the third and the fourth are *"high"* and the MR factors values are all of them *"high"* and, while the last 13 values of LR are all have *"high"* values.

Then the trust percentage will be:

$v_{f_{max}} = 8$, HR= $v_{f_{max}} \times 10^{v_{f_{max}}}$, MR= $v_{f_{max}} \times 10^{v_{f_{max}}-2}$, LR=$v_f \times 10^{v_{f_{max}}-5}$, v_f Ordered Descending

HR MR LR

Trust Percent=($8*10^{\mathbf{8}}+8*10^{7}+6*10^{6}+9*10^{\mathbf{6}}+8*10^{5}+8*10^{4}+8*10^{3}+9*10^{\mathbf{3}}+9*10^{2}+9*10^{1}+9*10^{0}$
$+9*10^{-1}+9*10^{-2}+9*10^{-3}+9*10^{-4}+9*10^{-5}+9*10^{-6}+9*10^{-7}+9*10^{-8}+9*10^{-9}$)/$10^{8}$

LR $10^{v_{f_{max}}}$

= 0.895898≈ **89.59%**

This result 89.59% indicates that our model generates a *high* trust value in terms of accuracy of information.

Chapter Six

CTA Agent Implementation

Chapter 6: CTA Agent Implementation

6.1 Introduction

In this chapter we will introduce how CTA agent works with factor analysis agents by decomposing content trust measurement process into one level of factor calculation tasks whereby each of these tasks allocated to an agent. We will be building and running these agents using *Agelts* and its programming interface *Java Aglet API* and running the Aglet Framework using ***Tahiti*** server to be prepared for accepting agents.

6.2 CTA Agent Algorithm And Programming

6.2.1 CTA Agent Algorithm

As we mentioned in chapter 4 the factor analysis agents are getting prepared to be dispatched in our environment after their knowledge are acceptable to answer CTA agent. Then the CTA agent is ready to be created and dispatched to our agents environment. The CTA agent starts sending different type of message passing techniques to factor analysis agents requesting for a value of specific webpage content URL. The types of message passing depend on factor classification that we showed it in chapter 5 as following:

TABLE 6.1:THE MESSAGE PASSING TYPES USED BY CTA AGENT FOR EACH CLASSIFICATION OF FACTOR ANALYSIS AGENTS.

Factor Classification	**CTA agent message passing type**
Low Rational	Future-Type
Medium Rational	Future-Type
High Rational	Now-Type

The *future-type* message passing is asynchronous and does not block the current execution. The method returns a `FutureReply` object, which can be used to obtain the result or wait for it later. Example1 shows how we can exploit the waiting time for CTA agent during requesting from *low rational LR* factor agent value, where exploitation was into requesting at most another 4 messages to *high rational HR* factor agents for their values until the result of low rational

factor agent is ready. We can know how many *high rational HR* factor agents are executed from waiting time exploitation from `num_HR` counter.

Example1:

```
FutureReply future =
         proxy.sendAsyncMessage(new Message("LR_url"));
     int num_HR = 4;
     // do high rational messages requests at most 4 times while waiting for
the result of first LR message.
     while(future.isAvailable() == false && num_HR-- >0) {
String answer = proxy.sendMessage(new Message("HR_url"));
         System.out.println(answer);

     }
     System.out.println( (String)future.getReply() );
```

A *now-type* message is synchronous and blocks the current execution until the receiver has completed the handling of the message. In the example 1 the following sub-code section is represent now-type messaging:

```
         String answer = proxy.sendMessage(new Message("HR_url"));
         System.out.println(answer);
```

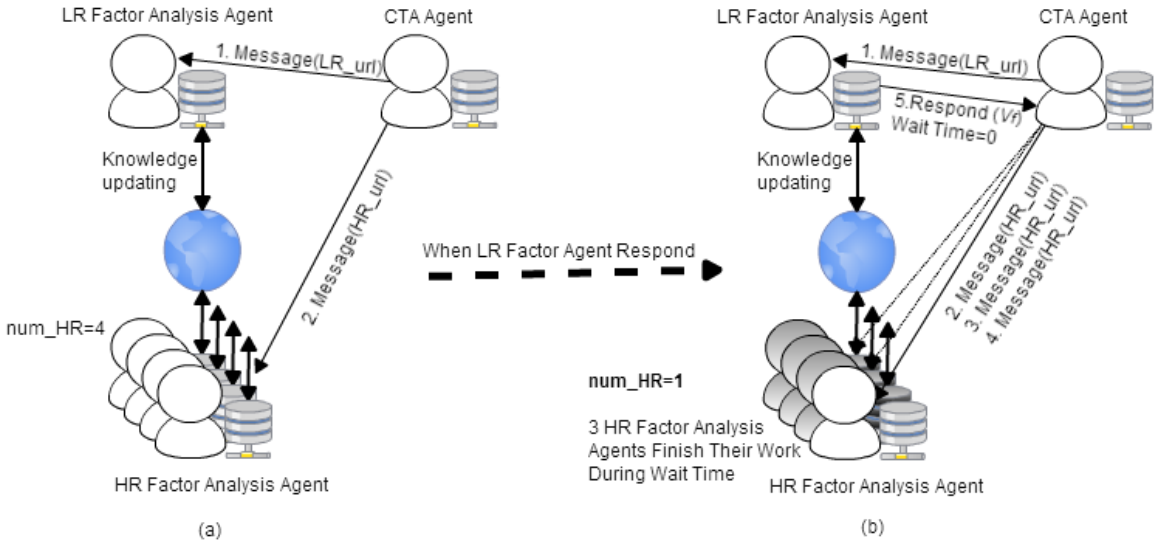


FIGURE6.1: A) CTA AGENT SEND FUTURE-TYPE(ASYNCHRONOUS) TO LR FACTOR AGENT EXPLOITS WAIT TIME BY SENDING AS IT CAN NOW-TYPE MESSAGES TO HR FACTOR AGENTS . B) SHOWS THAT A THREE NOW-TYPE MESSAGES TO HR FACTOR AGENTS ARE EXPLOITED WHEN LR FACTOR AGENT RESPOND.

The figure 6.1 shows how the exploitation of waiting time that generates from Future-type message passes through and uses as much as we can the Now-type message passing during the "wait time." In figure 6.1 three of four high rational factor agents are provide their values while the fourth could be executed in anther waiting time during another Future-type message.

Why is it that each factor classification is associated with specific type of message passing? The answer depends on the importance of each classification since high rational is the highest and the most affected on the final content trust value, so we used Now-type message passing with high rational. On the other hand, we used Future-type message passing with MR and LR factor agents because we can exploit the waiting time of all MR and LR factor agents responding to CTA agent in executing and messaging HR factor agents (that have Now-type messaging).

The receiver agent (could be factor agent or CTA) has to define its `handleMessage(Message msg)` method to handle incoming messages. In `thehandleMessage()` method, a message object is passed as an argument and can be used to perform the operation according to the kind of message. In the method, we have to return a `boolean` value indicating whether it has been handled or not. Example 2 explains how any factor analysis handle requested messages that coming to it from CTA agent.

Example2:

```
public boolean handleMessage(Message msg) {
    if (msg.sameKind("vf1")) {
        System.out.println("vf1");
        return true; // i know this message...
    }
    return false; // false, otherwise
}
```

If it returns false, the CTA agent that represents the sender of the request message receives a `NotHandledException` and thus knows that the message has not been handled.

```
Future future = proxy.sendAsyncMessage();
...
try {
    Object reply = future.getReply();
} catch (NotHandledException ex) {
    // the receiver didn't handled the message
} catch (MessageException ex) {
    // an exception has been thrown in the receiver's handleMessage()
    System.out.println(ex.getException());
}
```

Finally, the overall communication process between CTA agent and Factor analysis agents stop when at least HR and MR factor agents completely respond to CTA agent request since LR factor agents don't have critical effects on content trust measure if most of their responds not handled. Then the CTA agent starts the calculation based on collective factor agents' values using our model as we explained in chapter 5.

6.2.2 CTA Agent Programming

To program our agent algorithm, we choose Java language, since we choose to run our agent on aglet that have a programming interface *Java Aglet API*. We will talk about aglet with more details in next section. Another reason for choosing Java is due to the existence of many libraries or services that help us to capture factors values from the web easily.

6.3 Agents Running

6.3.1 Introduction To Aglet

We start building the agent and running it using *Agelts* and its programming interface *Java Aglet API*. The Java Aglet API (J-AAPI) [19] is a proposed industry standard for Javabased mobile agents. J-AAPI was developed by a research team at the IBM Tokyo Research Laboratory in Japan in response to a call for a uniform platform for mobile agents in heterogeneous environments such as the Internet [16].

Aglets provide in a unique and powerful way a new and viable paradigm that unifies:

- Mobile and stationary objects.
- Object-passing, message-passing, and data-passing.
- Autonomous and passive objects.
- Asynchronous and synchronous processing.
- Local and remote objects.
- Disconnected and connected operations, and
- Parallel and sequential execution [19].

Our works will use this framework to running our agent as stand-alone agent, we looking for future works to develop this agent to cooperate with other agents and more. (See chapter 6).

6.3.2 Running Aglet Framework (Tahiti)

We will show in this section how we can run the Aglet Framework using ***Tahiti*** server to be prepared for accepting agents. *Tahiti* is an application program that runs as an agent server. You can run multiple servers *(Tahiti)* on a single computer by assigning them different port

numbers. Tahiti provides a user interface for monitoring, creating, dispatching, and disposing of agents and for setting the agent access privileges for the agent server.

Java Development Kit (JDK) must be installed in the machine that we want to run *Tahiti* server. Let's indicate for the directory path under where we installed JDK *$(JDK)*, and *$(ASDK)* indicates the directory path beneath the ASDK (Agent Software Development Kit), for example in our work the paths are:

$(ASDK) ➔ `C:\aglets-2.5-gamma\` (Win95, 98, NT, OS/2)

And *$(JDK)* ➔ `c:\jdk1.7.0_05\` (Win95, 98, NT, OS/2)

Then we started setting up environment variables before using startup script (batch file), including AGLET_HOME and JDK_HOME. The startup script file can be found at the directory named ***bin*** under where we installed ASDK. The name of the startup script file is,

- $(ASDK)`\bin\agletsd.bat` (Win95, 98, NT)

The way to set environment variables for ASDK is to insert the following two lines at the beginning of the script file. For example,

- `set JDK_HOME=$(JDK)`
- `set AGLET_HOME=$(ASDK)`

We can start up Tahiti by creating a command prompt window and start the daemon by executing a script in the aglets package:

```
$(ASDK)\bin\agletsd
```

This daemon will listen for agents on the default port number, 4434. We can create another command prompt window and start another daemon here:

```
$(ASDK)\bin\agletsd -port 2000
```

This daemon will listen for agents on another **-port** in this case port number 2000. You can actually choose any port that is not occupied by some other application and its operating system. The daemon also makes Tahiti appear on your screen. Tahiti is a desktop interface for managing running agents.(See figure 6.2).

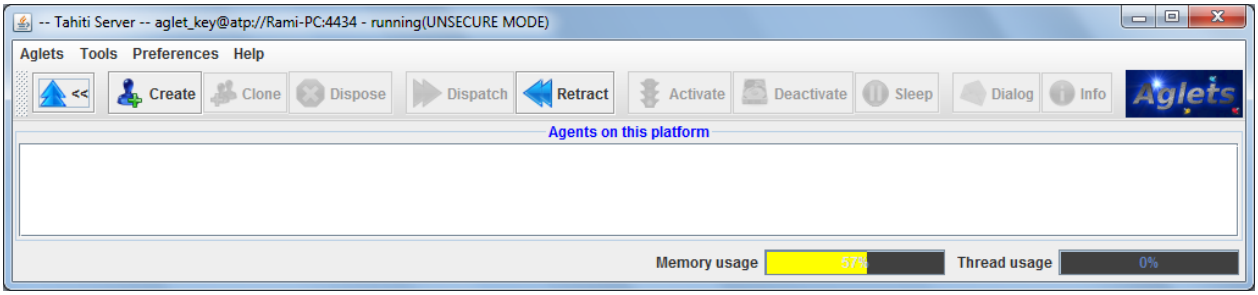


FIGURE62 .:TAHITI SERVER RUNNING.

6.3.3 Running Content Trust Agent On Tahiti

The first thing to be observed in running agent is the compliance of its architecture. The figure 6.3 shows this architecture. We can add any libraries or packages we need but the statement "`import com.ibm.aglet.*;`" must always exist. The function `run()` is equivalent to `main()` function in Java. We can write in it the main code of agent. Finally the agent class must be `extends` the `Aglet` class to work properly in Tahiti server. *(Note: the whole code in Appendix B)*

```
...
import com.ibm.aglet.*;
...
public class MyAglet extends Aglet {
    ...
      public void run() {
      ...
      }

}
```

FIGURE6.3:AGENT ARCHITECTURE IN AGLET.

After that we can compile the agent class "`MyAglet`" to be ready to run on Tahiti server, any related class files to the main class of agent must be in the same package before Tahiti server import them.

The following step is creating a new agent in Tahiti server as shown in figure 6.4. The new agent window appears asking to fill the package of classes and the main class name and the path of the package. (See figure 6.5) then we press "Add to the list " button then "Create".

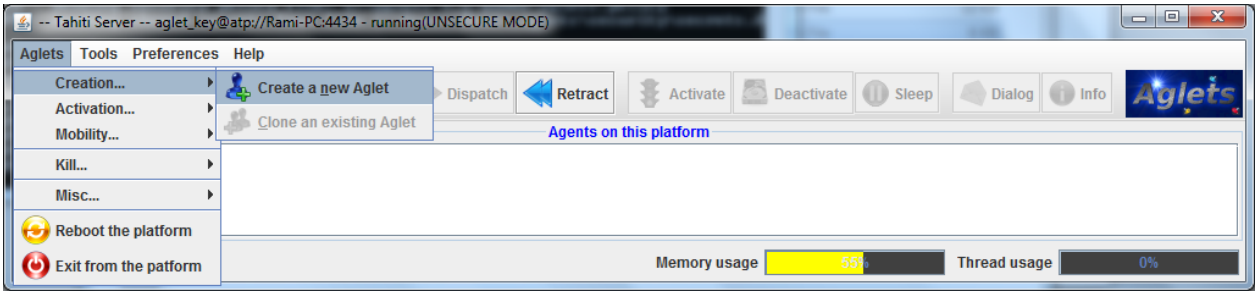


FIGURE 6.4 : CREATE A NEW AGENT ON TAHITI SERVER.

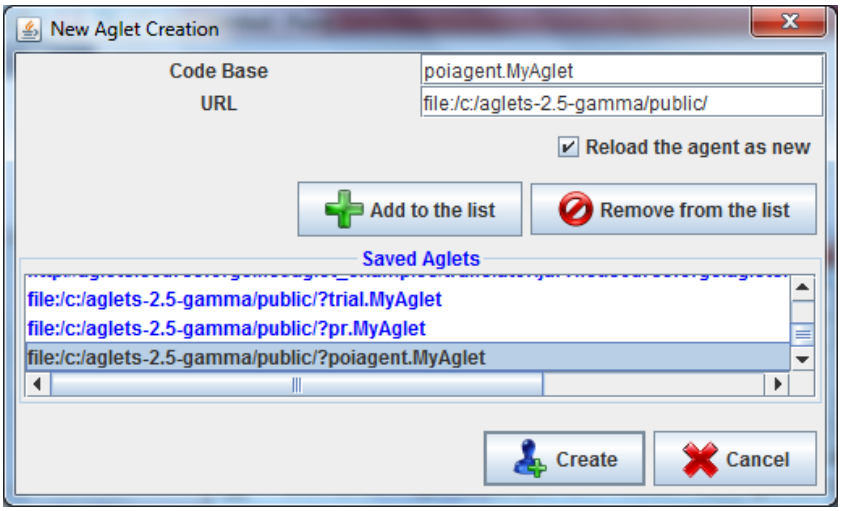


FIGURE 6.5: CREATE A NEW AGENT WINDOW.

The agent will run immediately, starting its calculation and works, figure 6.6 shows the trust agent in running state inside Tahiti server, while figure 6.7 shows the results of this agent in command promote and store them in database file to be analyzed later in "results analysis" section.

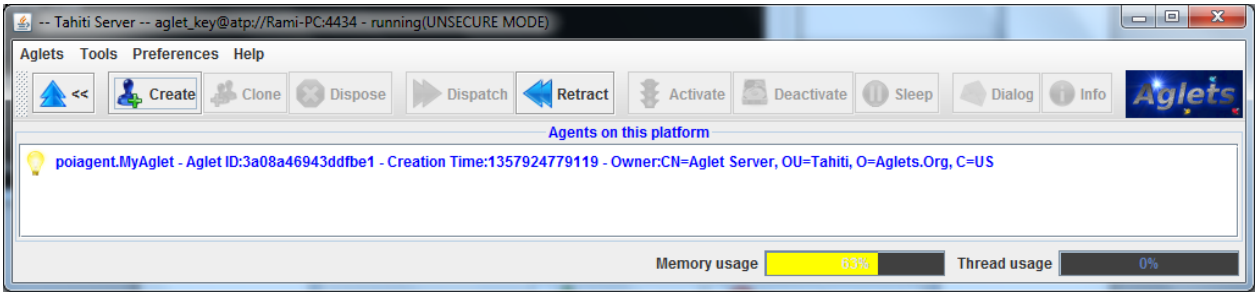


FIGURE6.6: CONTENT TRUST AGENT RUNNING IN TAHITI

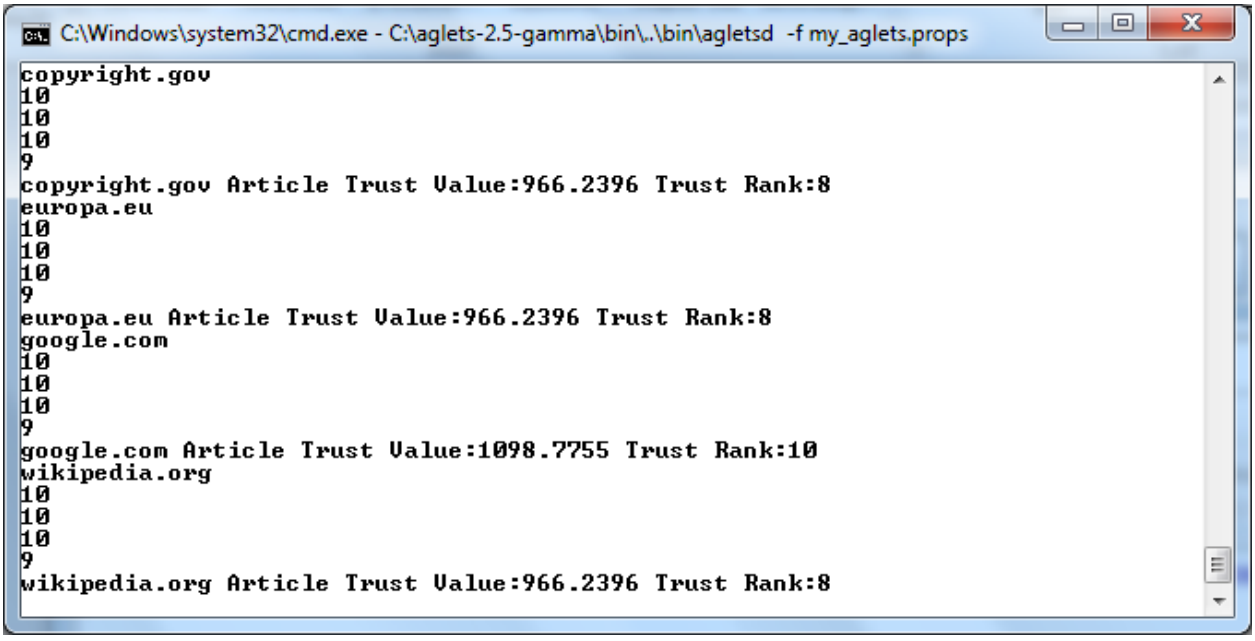


FIGURE6.7 :CONTENT TRUST AGENT RESULTS IN CMD.

Chapter Seven

Model Evaluation

Chapter 7: Model Evaluation

7.1 Introduction

In this chapter, we want to explain what and how our research is evaluated? Defining a confidence interval, hotelling's T^2 and how we use them in order to determine the most websites that having maximum errors. Finally comes the task of analyzing the result of evaluation experiment and concluding the results of our work.

7.2 Evaluation Method

We want to evaluate the outputs of the trust model of our multi-agent system, which generate a trust value and see how close that value *(accuracy)* is to human perception *(from expert users)*. One method of evaluating the accuracy of MAS output is attained by comparing it with expert opinions *(to keep following human perception especially form expert users)* to achieve that two main steps: first, empirical evaluation experiment done through a set of expert users on several web information resources (articles). Second we compare the results from CTA agent with the results from the respondents and see how much CTA agent outputs deviate from experts' opinion *(that represents the benchmark in evaluation)*, to achieve that we will:

1- Show how many CTA agent answers are within the *confidence interval* of average respondent answers for each information source. We use this method for two purposes:

   a. Calculate *the accuracy* of trust model in MAS where the accuracy means the percent of number of webpages within confidence interval to overall web-pages count.

   b. Studying *unmatched webpages* by our model by analyzing how the factors in these unmatched web-pages are evaluated from experts' point of view and CTA agent calculations.

2- Applying hotelling's T^2 statistics analysis in *appendix F* by testing whether the CTA agent answers vector μ_0 is a plausible value for the respondents population answers mean vector μ.

3- Applying hotelling's T^2 statistics analysis after *excluding* the unmatched webpages (*discussed in appendix F*).

4- Studying and analyzing demographic of experts how affects the accuracy of trust for different respondents gender, job experience and majors.

- *Confidence Interval:*

The *confidence interval* is a term used in inferential statistics that measures the probability that a population parameter will fall between two set values. The confidence interval can take any number of probabilities, with the most common being 95% or 99%. In other words, a confidence interval is the probability that a value will fall between an upper and lower bound of a probability distribution.

For example, a machine is supposed to fill "2-Liter" bottles of Pepsi. To see if the machine is working properly, we randomly select 100 bottles recently filled by that machine, and find that the average amount of Pepsi is 1.985 liters. Can we conclude that the machine is not working properly?

No! By simply reporting that $\bar{x} = 1.985$ liters, we are neglecting the fact that the amount of Pepsi varies from bottle to bottle and that the value of the sample mean depends on the luck of the draw. It is possible that a value as low as 1.985 is within the range of natural variability for $\bar{X}$, even if the average amount for all bottles is in fact $\mu = 2$ liters. Let us suppose we know from past experience that the amounts of Pepsi in bottles filled by the machine have a standard deviation of $\sigma = 0.05$ liters, since n = 100, we can assume (using the Central Limit Theorem) that $\bar{X}$ is normally distributed with mean μ (unknown) and *standard error*:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 0.005$$

So the probability is about 0.95 that μ will be within 0.01 liters of $\bar{X}$. Thus, the interval $\bar{X} \pm 0.01$will contain μ with a probability of 0.95. In general, the interval $\bar{X} \pm 2\frac{\sigma}{\sqrt{n}}$ will contain μ with probability about 0.95. Therefore, the interval provides (approximately) a 95% confidence interval for μ. From the Empirical Rule, the probability is about 95% that $\bar{X}$ will be within two standard errors of its mean.

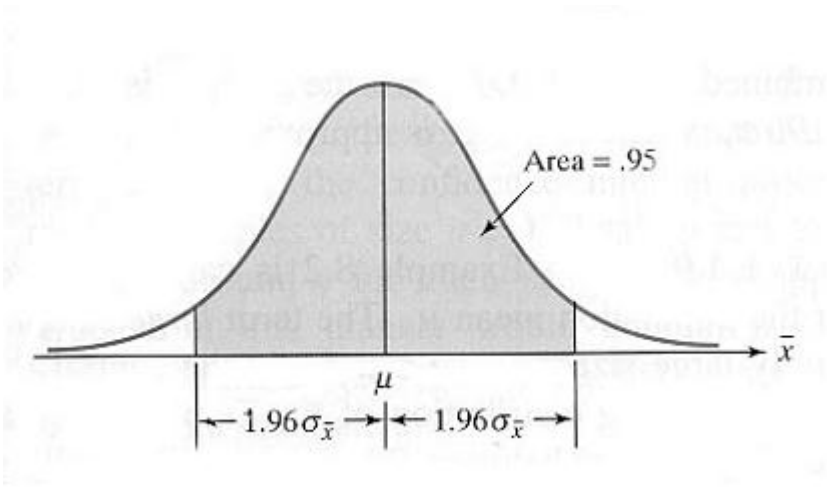


FIGURE 7.1: 95% CONFIDENCE INTERVAL FOR M.

- *Hotelling's T^2:*

We want to discover the websites that have maximum errors and how these errors affect overall group accuracy. Hotelling's T^2 test is used to achieve that for further information about hotelling's T^2 test. For further details, see the appendix F that shows whether the CTA agent answers vector μ_0 (*P x 1* dimension where *P*: is number of webpages) is a plausible value for the respondent population answers mean vector μ before and after excluding high error websites.

- *Evaluation of Model Design:*

We evaluate our multi-agent system using an online questionnaire (*See Appendix D*) which is done through a 44 *expert- user sample and* 30 web information resources (articles and information), 4 respondents are excluded because of their unreasonable answers and take a high exclusion value. The 40 other respondents are divided into two different groups *(running sets)* each containing 20 experts on 15 web information resources. We aim to see if with two different groups the model still provides reasonable answers or not.

The questionnaire will show a variety of website articles, each of which contains the information that we want to judge, where each person is required to read them and show the percentage of trust. The information sources we selected in each group have different articles but they implicitly have the same level trust factors values, for example the first information source URL that exists in group1 questionnaire is different in its content from the first information source URL in group2 but both of them have *"High Authority"* and *"Low Appearance"*. This variety of factor selection in webpage articles and different factor values make us insure that the model work in all factors cases.

Expert users are people dealing with web at least 7 hours in a week, besides they have a good education and expert in their work. The main aim for selecting experts is to make sure if our CTA agent model achieves good and reasonable results according to expert users answers, so we select in our evaluation 40 lectures from Palestine Polytechnic University (PPU) to be expert users in our experiment.

7.3 Experiment Results Analysis

We will first show the factor analysis agents returned their values for the requesting messages from CTA agent. Then we will analyze demographic information of the users of survey and explain how CTA agent answer is related to this information. Afterwards, we compare the average respondents answers to our CTA agent for each webpage article.

7.3.1 Factor Analysis Agents Results

The factor analysis agents start their work for each request from CTA agent as cooperation process between them. Each one of them has its own knowledge (database) which is updated continuously. Besides, these updates could take many days for most factor agents to generate acceptable knowledge e.g. appearance, age and update… etc. The table 7.1 and table 7.2 shows the *value* (v_f) that returned from factor agents for group1 and group2 respectively, where selected information source URLs abbreviated in format of to Ws_ng_m where n is webpage number and m indicates for the group, for example Ws1g1 means the first webpage that appears to group1 *Appendix E* shows an abbreviation for URL link for the 15 website article and information were we selected for each group.

Running 8 factor agents in our multi-agents system for the following reasons:

- Since some factors are hard to be implement or identified as we mentioned in previous chapters. E.g. *referencing*
- Ignoring factors lies on low rational factor area will not affect critically on our system, as we will see in next section (see figure 7.5). E.g. *Links Inconsistency, Evaluating and Comments, Number Of Visitors, …etc.*

TABLE 7.1:FACTOR ANALYSIS AGENTS RETURNED VALUES FOR FIRST GROUP.

	vf1	*vf2*	*vf3*	*vf4*	*vf5*	*vf6*	*vf7*	*vf8*
Ws1g1	8	Categorized URL	7	2070	1	7	Minimal Risk	2
Ws2g1	3	Uncategorized URL	0	18421	1	8	Unverified	9
Ws3g1	2	Categorized URL	1	117664	1	10	High Risk	1
Ws4g1	3	Uncategorized URL	0	232342	5	9	Medium Risk	3
Ws5g1	3	Categorized URL	2	162346	1	9	Medium Risk	3
Ws6g1	6	Categorized URL	3	529220	1	9	High Risk	2
Ws7g1	7	Categorized URL	5	796927	10	6	Unverified	8
Ws8g1	8	Categorized URL	4	74677	1	5	Medium Risk	5
Ws9g1	3	Categorized URL	1	306686	1	9	High Risk	4
Ws10g1	7	Categorized URL	6	653466	1	10	Minimal Risk	5
Ws11g1	7	Uncategorized URL	4	395839	8	9	Unverified	4
Ws12g1	6	Categorized URL	5	47491	1	9	Minimal Risk	4
Ws13g1	4	Uncategorized URL	3	327278	10	5	Unverified	7
Ws14g1	6	Categorized URL	5	73963	1	9	Minimal Risk	3
Ws15g1	6	Categorized URL	2	655032	10	6	High Risk	9

TABLE 7.2: FACTOR ANALYSIS AGENTS RETURNED VALUES FOR SECOND GROUP.

	vf1	*vf2*	*vf3*	*vf4*	*vf5*	*vf6*	*vf7*	*vf8*
Ws1g2	7	Categorized URL	1	546114	2	9	Minimal Risk	1
Ws2g2	5	Categorized URL	2	40535	1	7	High Risk	1
Ws3g2	7	Uncategorized URL	3	380192	1	9	Medium Risk	2
Ws4g2	8	Categorized URL	8	1840	1	6	Minimal Risk	2
Ws5g2	7	Uncategorized URL	0	839100	10	7	Unverified	1
Ws6g2	8	Categorized URL	5	32522	1	9	Minimal Risk	2
Ws7g2	3	Categorized URL	3	189145	10	8	Minimal Risk	2
Ws8g2	3	Categorized URL	7	1507	1	8	Minimal Risk	1
Ws9g2	7	Uncategorized URL	1	522811	7	9	Unverified	2
Ws10g2	6	Categorized URL	7	3514	1	9	Minimal Risk	2
Ws11g2	3	Uncategorized URL	1	973903	1	7	Medium Risk	2
Ws12g2	7	Categorized URL	3	724144	10	9	Minimal Risk	4
Ws13g2	3	Uncategorized URL	1	47495	4	1	Unverified	3
Ws14g2	6	Uncategorized URL	1	477378	1	9	Medium Risk	3
Ws15g2	6	Categorized URL	2	655032	10	6	High Risk	9

Where:

- ***vf1*: the value from "Appearance" agent.**
- ***vf2*: the value from "Specialization" agent.**
- ***vf3*: the value from "Authority" agent.**
- ***vf4*: the value from "Popularity" agent.**
- ***vf5*: the value from "Age and Update" agent.**
- ***vf6*: the value from "Methodology of writing" agent.**
- ***vf7*: the value from "Security" agent.**
- ***vf8*: the value from "Locality" agent.**

7.3.2 CTA agent and Respondents Information Analysis

- *Confidence Interval Analysis:*

To insure that our CTA agent answers vector μ_0 was rational we compare it with respondents average answers vector μ. The figure 7.3 shows how CTA agent results appear within the range of confidence interval of respondents answers $[\mu - t_{n-1}\frac{\sigma}{\sqrt{n}}, \mu + t_{n-1}\frac{\sigma}{\sqrt{n}}]$, where *n* is population size, σ is standard deviation of population and $t_{n-1} = t_{20-1} = t_{19}$=2.093 from *T- Distribution Table* as appeared in figure 7.2.

Pr / df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.

FIGURE 7. 2: T-DISTRIBUTION TABLE FOR EACH NUMBER OF POPULATION

In figure 7.3, we showed whether the result from CTA agent was within confidence interval of average respondents answers in group1 for each webpage. *Ws10g1, Ws12g1 and Ws13g1* are the only websites that CTA agent answer does not belong to their confidence interval; this is due to *referencing* factor is missing in our CTA agent calculation this factor appears very clearly in this website article in *Ws13g1*. In *Ws10g1and Ws12g1* case we notice that some factors play their role in our model implicitly away from the eyes of respondents like: authority, security, age and update … etc. when these factors have high values, they affect our model by exceeding the CTA agent answer more positively.

Calculating the accuracy of our model for group1 respondents:

$$Accuracy_{Group1} = \frac{Number\ of\ webpages\ within\ confidence\ interval}{Overall\ webpages\ count} \times 100\%$$

$$= \frac{12}{15} \times 100\% = 80.0\%$$

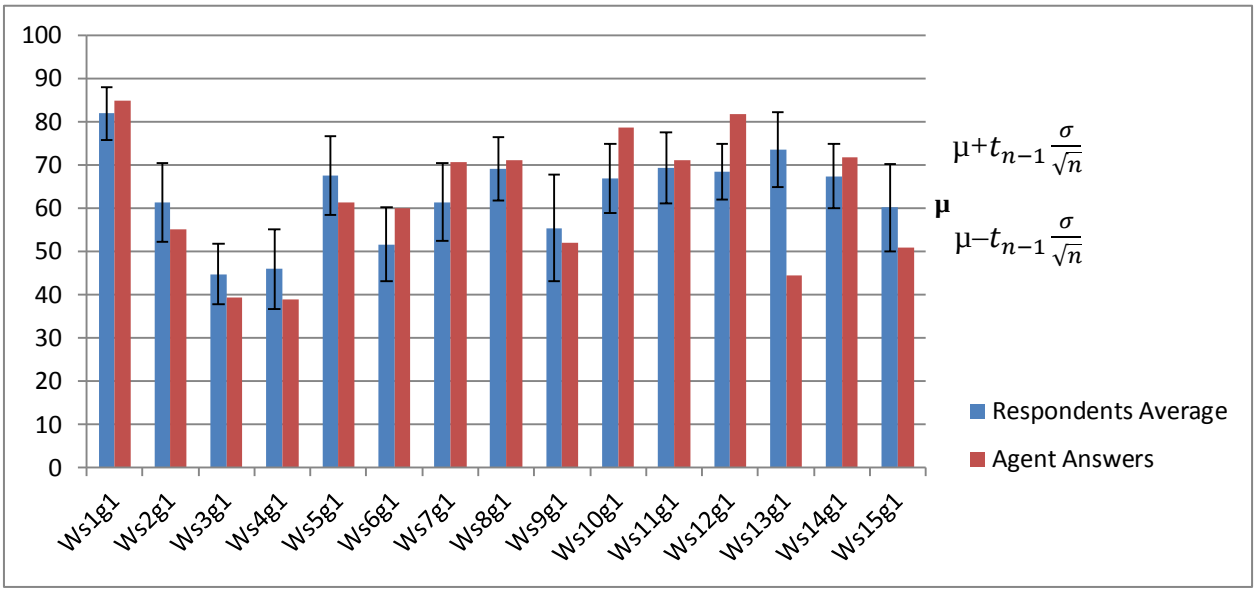


FIGURE 7. 3: RESPONDENTS AVERAGE ANSWERS VS. CTA AGENT ANSWERS WITHIN [$\mu-t_{n-1}\frac{\sigma}{\sqrt{n}}$, $\mu+t_{n-1}\frac{\sigma}{\sqrt{n}}$]. (GROUP 1)

Figure 7.4 shows CTA agent answer according to respondents results within **[$\mu-t_{n-1}\frac{\sigma}{\sqrt{n}}$, $\mu+t_{n-1}\frac{\sigma}{\sqrt{n}}$]** for the second group, Ws4g2, Ws11g2, Ws12g2 and Ws15g2 are exceed the standard deviation interval. We believe that two things could improve these results, first by increasing the number of factors implemented in the CTA agent, Second is more study about the relation between these factors and how this could affect each other to produce appropriate trust judgment. For example, Ws15g2 has popularity-locality relation problem where what is popular in your country could not

be popular in other countries. We study some of these relations in group2 and show how it could improve the results in section 7.3.4.

If we want to calculate the accuracy of our model for group2 respondents without handling any relation between factor agents, then:

$$Accuracy_{Group2} = \frac{Number\ of\ webpages\ within\ confidence\ interval}{Overall\ webpages\ count} \times 100\%$$

$$= \frac{11}{15} \times 100\% = 73.3\%^{\dagger}$$

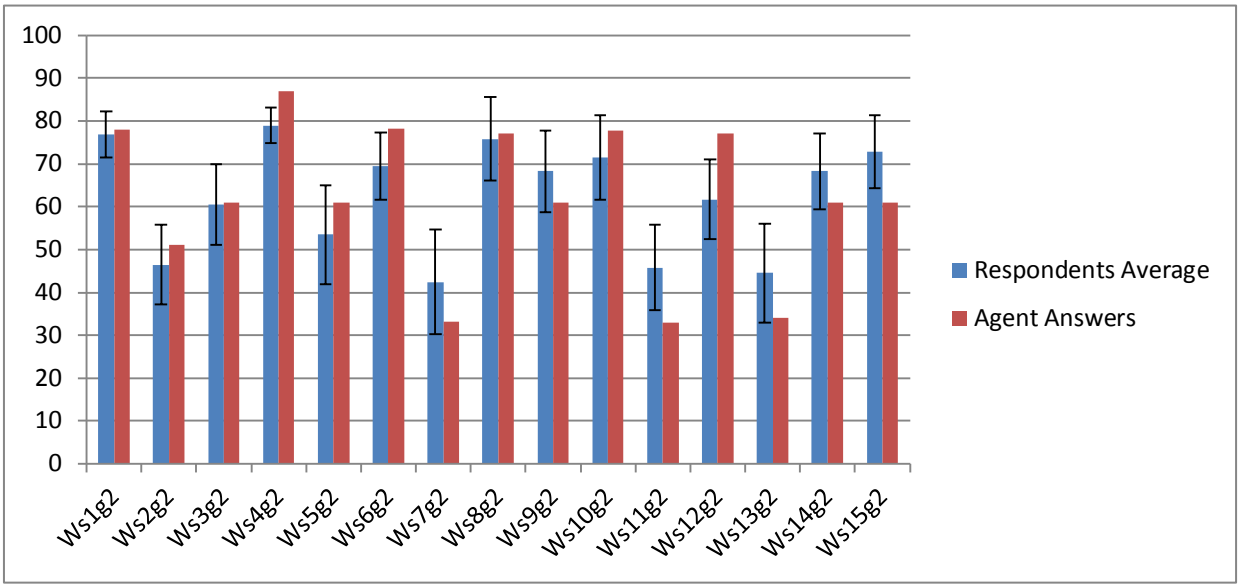


FIGURE 7. 4: RESPONDENTS AVERAGE ANSWERS VS. CTA AGENTS ANSWERS WITHIN $[\mu - t_{n-1}\frac{\sigma}{\sqrt{n}}, \mu + t_{n-1}\frac{\sigma}{\sqrt{n}}]$. (GROUP 2)

We can take the overall accuracy of our model as $\mathbf{Trust_{Accuracy}} = \frac{80+73.3}{2} = \mathbf{76.65\%}$, we calculate respondents trust percentage about what they read in 15 websites as represented precisely in table 7.3 for group1 and table 7.4 for group2. The CTA agent starts automatically to calculate its value for each article and compare its value with all respondents' answers.

[†] This value was improved to 80% after we implement the *Popularity and Locality* relation on group2 that makes Ws15g2 within its confidence interval, see section 7.3.4 *Factor Analysis Agents Dependences*

TABLE 7.3: RESPONDENTS RESULTS COMPARED TO CTA AGENT CALCULATION (FIRST GROUP).

For Group 1	Ws1g1	Ws2g1	Ws3g1	Ws4g1	Ws5g1	Ws6g1	Ws7g1	Ws8g1	Ws9g1	Ws10g1	Ws11g1	Ws12g1	Ws13g1	Ws14g1	Ws15g1
Respondent1	92	70	50	40	95	55	97	85	90	95	95	86	87	90	94
Respondent2	82	30	30	50	50	40	50	60	60	65	60	50	50	70	40
Respondent3	81	50	30	50	50	40	30	90	70	70	80	50	70	40	46
Respondent4	95	40	20	10	80	70	70	80	75	70	80	80	60	80	85
Respondent5	100	60	35	50	75	50	90	80	90	85	80	75	85	90	50
Respondent6	60	70	50	75	80	40	30	65	90	85	50	50	90	85	80
Respondent7	90	90	50	30	80	70	50	90	10	80	100	90	100	40	20
Respondent8	70	65	53	75	80	50	50	80	75	60	65	80	65	50	75
Respondent9	70	80	55	40	95	90	65	55	20	65	85	65	88	60	40
Respondent10	99	80	70	60	80	40	60	70	80	80	30	80	30	80	70
Respondent11	80	43	26	50	33	27	37	57	30	44	70	90	55	66	34
Respondent12	79	87	60	60	40	40	80	70	70	40	40	60	50	60	70
Respondent13	85	80	30	40	70	70	70	70	45	40	80	60	70	80	60
Respondent14	90	55	55	10	40	24	40	50	25	80	75	55	60	55	85
Respondent15	50	20	66	85	45	67	88	78	27	43	78	78	87	67	66
Respondent16	75	65	50	40	70	60	75	50	75	50	60	66	78	80	40
Respondent17	70	65	60	45	70	50	60	30	40	60	65	78	85	60	70
Respondent18	82	40	45	40	50	70	55	60	50	60	55	65	80	70	60
Respondent19	95	80	20	50	90	20	60	80	20	90	60	50	100	45	30
Respondent20	93	55	40	20	80	60	70	82	66	75	80	60	81	80	88
Respondents Average	**81.9**	**61.25**	**44.75**	**46**	**67.65**	**51.65**	**61.35**	**69.1**	**55.4**	**66.85**	**69.4**	**68.4**	**73.55**	**67.4**	**60.15**
Agent Answers	**84.872**	**55.0817**	**39.2911**	**38.9955**	**61.3115**	**60.0012**	**70.5667**	**71.1055**	**52.0811**	**78.662**	**71.0317**	**81.742**	**44.3887**	**71.742**	**50.9521**
Absolute Difference	2.972001	6.1683	5.4589	7.0045	6.3385	8.3512	9.2167	2.0055	3.3189	11.812	1.6317	13.342	29.1613	4.342	9.1979
Standard Deviation σ	13.20247	19.48245	15.02585	19.7084	19.41046	18.15801	19.22245	15.62016	26.3846	17.10117	17.48202	13.85793	18.4033	15.836	21.54866
Confidence	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%
Critical t	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024
Standard Error	2.952162	4.356408	3.359883	4.406932	4.34031	4.060253	4.298271	3.492774	5.899777	3.823938	3.909099	3.098726	4.115104	3.541038	4.818427
Lower Limit	75.72105	52.13193	37.71768	36.77618	58.56563	43.15179	52.35362	61.78954	43.05162	58.84641	61.21816	61.91429	64.93699	59.98852	50.06492
Upper Limit	88.07895	70.36807	51.78232	55.22382	76.73437	60.14821	70.34638	76.41046	67.74838	74.85359	77.58184	74.88571	82.16301	74.81148	70.23508

: Missed webpages in our results

TABLE 7.4: RESPONDENTS RESULTS COMPARED TO CTA AGENT CALCULATION (SECOND GROUP).

	Ws1g2	Ws2g2	Ws3g2	Ws4g2	Ws5g2	Ws6g2	Ws7g2	Ws8g2	Ws9g2	Ws10g2	Ws11g2	Ws12g2	Ws13g2	Ws14g2	Ws15g2
Respondent1	75	35	70	75	78	83	25	65	70	86	40	80	20	60	40
Respondent2	93	60	41	90	0	50	20	20	70	85	10	30	50	70	86
Respondent3	90	60	80	85	30	80	0	100	60	100	20	60	90	90	70
Respondent4	82	30	50	65	27	22	78	45	15	67	47	44	34	67	84
Respondent5	80	70	90	70	80	90	80	90	60	85	70	70	60	90	90
Respondent6	60	70	50	70	60	70	70	60	80	70	70	50	60	70	50
Respondent7	60	50	60	80	50	70	50	80	75	85	50	85	50	90	80
Respondent8	65	70	85	90	90	90	85	90	75	90	85	70	70	90	90
Respondent9	70	75	80	90	50	85	75	50	20	20	30	75	20	50	95
Respondent10	80	60	65	85	80	60	55	73	82	63	84	80	90	74	90
Respondent11	78	50	75	60	65	70	70	80	90	70	80	90	50	90	80
Respondent12	90	10	30	90	30	80	30	90	80	90	30	50	40	30	90
Respondent13	86	45	65	70	60	50	45	85	50	45	45	55	70	20	50
Respondent14	77	70	80	80	20	60	35	95	70	40	30	30	40	60	70
Respondent15	60	30	30	75	40	70	30	60	80	70	40	60	10	75	70
Respondent16	65	20	40	75	60	60	25	85	85	40	45	40	60	80	60
Respondent17	95	25	40	80	60	70	10	100	90	70	40	35	10	65	80
Respondent18	70	30	75	80	80	85	20	74	65	92	35	88	23	65	45
Respondent19	91	40	30	90	30	60	25	100	85	70	30	55	20	65	90
Respondent20	70	30	75	80	80	85	20	74	65	92	35	88	23	65	45
Respondents Average	**76.85**	**46.5**	**60.55**	**79**	**53.5**	**69.5**	**42.4**	**75.8**	**68.35**	**71.5**	**45.8**	**61.75**	**44.5**	**68.3**	**72.75**
Agent Answers	**77.973**	**51.0921**	**60.9847**	**87.062**	**61.0787**	**77.142**	**33.0993**	**77.0823**	**60.98071**	**77.692**	**32.8515**	**77.0973**	**34.1417**	**60.9736**	**61.0682**
Absolute Difference	1.123001	4.5921	0.4347	8.062	7.5787	8.642	9.3007	1.2823	7.36929	6.192	12.9485	15.3473	10.3583	7.3264	11.6818
Standard Deviation σ	11.57254	19.67499	20.00651	8.973646	24.70137	16.76933	26.08569	20.8922	20.34511	21.19707	21.3211	19.98388	24.61172	19.04593	18.17423
Confidence	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%
Critical t	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024	2.093024
Standard Error	2.587699	4.399462	4.473592	2.006568	5.523395	3.749737	5.832937	4.671639	4.549306	4.739809	4.767544	4.46853	5.503348	4.258799	4.063881
Lower Limit	71.43388	37.29182	51.18666	74.8002	41.9394	61.65171	30.19152	66.02215	58.82819	61.57947	35.82142	52.39726	32.98136	59.38623	64.2442
Upper Limit	82.26612	55.70818	69.91334	83.1998	65.0606	77.34829	54.60848	85.57785	77.87181	81.42053	55.77858	71.10274	56.01864	77.21377	81.2558

: Missed webpages in our results

7.3.3 Analyze Demographic Information

- *Gender Analysis:*

As mention in previous section, 40 *expert users* were chosen to participate in the survey on 30-web information resources (articles and information) divided in two groups each containing 20 experts on 15 web information resources. The group 1 respondents include 5 females and 15 males. So, 75% of the respondents are male. On other hand the second group respondents include 7 females and 13 male, where 65% of respondents are male, *see figure 7.5*.

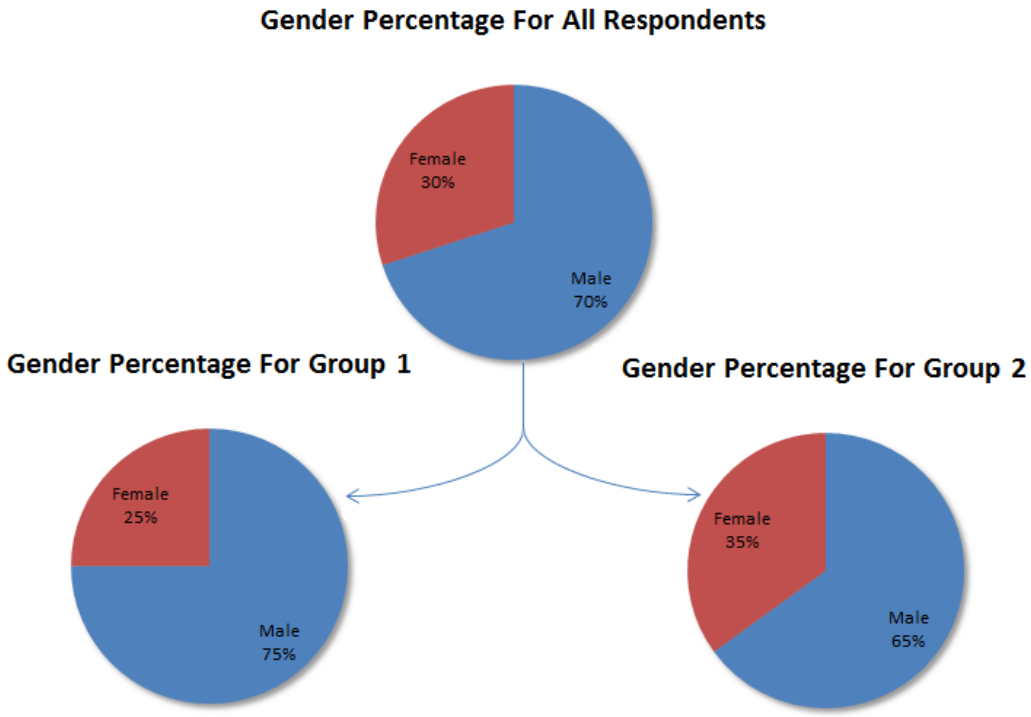


FIGURE 7. 5: RESPONDENTS GENDER DISTRIBUTION.

In group 1 there is no difference between the percentage of trust accuracy for "Male" gender and "Female" since the accuracy of trust for the "Male" was 73.3% and "Female" was 80%. The figure 7.6 show the comparison between average male respondents answers and CTA agent answers. It also checks if the average male respondents answers for each webpage article stays within confidence interval or not. The accuracy is calculated here by the percentage of how many average respondent answers don't exceeded the confidence interval to all 15 webpage articles. We see from the figure 7.6 that 4 webpage articles are out of their confidence intervals and they

are: Ws7g1, Ws10g1, Ws12g1 and Ws13g1. So the accuracy of male answers to CTA agent answer is:

$$Accuracy_{Male\ g1} = \frac{Number\ of\ webpages\ within\ confidence\ interval}{Overall\ webpages\ count} \times 100\%$$

$$= \frac{11}{15} \times 100\% = 73.3\%$$

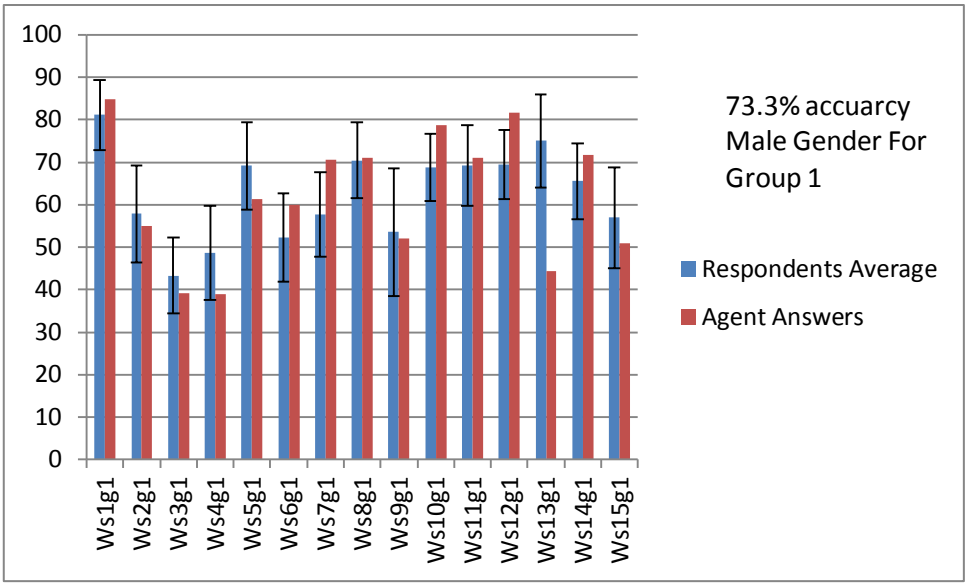


FIGURE 7.6: CTA AGENT ANSWER COMPARED TO AVERAGE MALE RESPONDENTS ANSWERS AND CONFIDENCE INTERVAL /GROUP1.

In figure 7.7 we notice that only 3 of average female respondents answers are exceeded their confidence interval and they are: Ws2g1, Ws12g1 and Ws13g1, so the trust value in this case is:

$$Accuracy_{Female\ g1} = \frac{Number\ of\ webpages\ within\ confidence\ interval}{Overall\ webpages\ count} \times 100\%$$

$$= \frac{12}{15} \times 100\% = 80\%$$

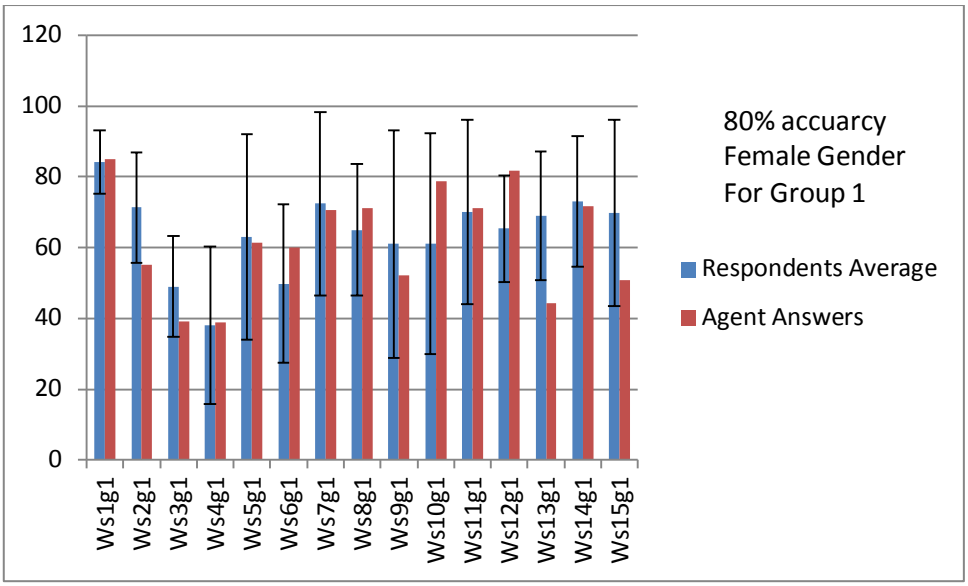


FIGURE 7.7: CTA AGENT ANSWER COMPARED TO AVERAGE FEMALE RESPONDENTS ANSWERS AND CONFIDENCE INTERVAL /GROUP1.

The figure 7.8 shows the male respondents results in group 2 and the webpages that lie out of confidence intervals are: Ws4g2, Ws7g2, Ws11g2, Ws13g2 and Ws14g2 while the accuracy was:

$$Accuracy_{Male\ g2} = \frac{Number\ of\ webpages\ within\ confidence\ interval}{Overall\ webpages\ count} \times 100\%$$

$$= \frac{10}{15} \times 100\% = 66.7\%$$

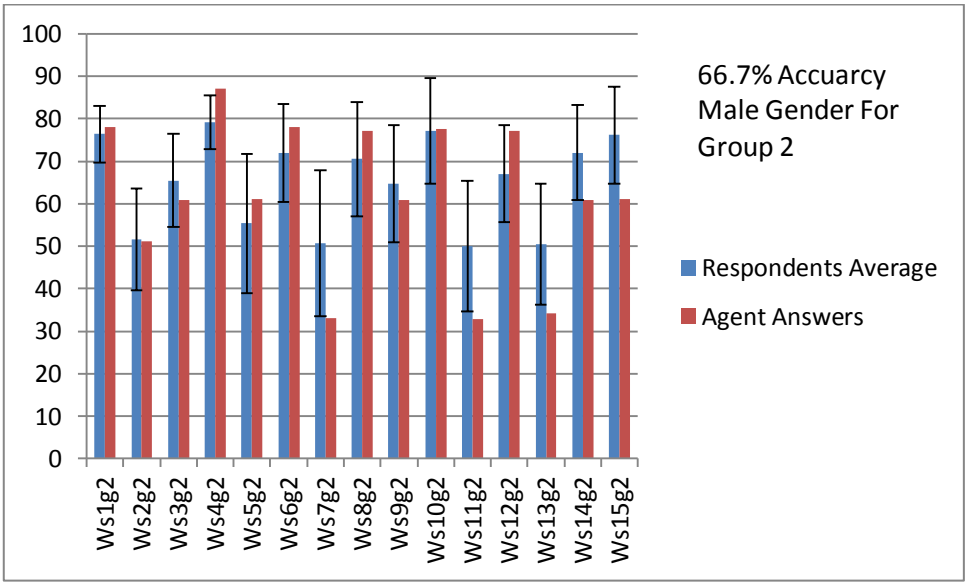


FIGURE 7. 8: CTA AGENT ANSWER COMPARED TO AVERAGE MALE RESPONDENTS ANSWERS AND CONFIDENCE INTERVAL /GROUP2.

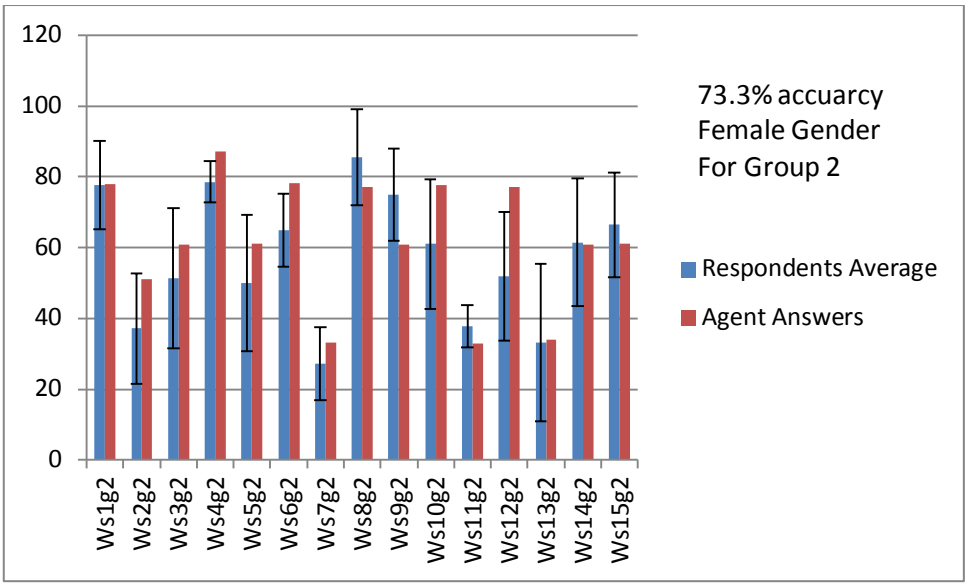


FIGURE 7. 9: CTA AGENT ANSWER COMPARED TO AVERAGE FEMALE RESPONDENTS ANSWERS AND CONFIDENCE INTERVAL /GROUP2.

The figure 7.9 shows how the results of female respondents look like, as well as the webpages that lie out of confidence intervals, namely: Ws4g2, Ws6g2, Ws9g2 and Ws12g2, we can derive the accuracy of the results for group2 female respondents by:

$$Accuracy_{Female\ g2} = \frac{Number\ of\ webpages\ within\ confidence\ interval}{Overall\ webpages\ count} \times 100\%$$

$$= \frac{11}{15} \times 100\% = 73.3\%$$

As a result, we notice that female respondents still provide relatively more accurate trust results that male respondents where female provide an accuracy with 80% while the males' rate is 73.3% in group1 and accuracy with 73.3% while males' rate is only 66.7% in group2. In addition, we can summarize the missed webpages according to respondents genders as following:

TABLE 7. 5 :MISSED WEBPAGES SUMMARY ACCORDING TO GENDER.

Group	Missed webpages	Gender
1	Ws7g1, Ws10g1, Ws12g1 and Ws13g1	Male
1	Ws2g1, Ws12g1 and Ws13g1	Female
2	Ws4g2, Ws7g2, Ws11g2, Ws13g2 and Ws14g2	Male
2	Ws4g2, Ws6g2, Ws9g2 and Ws12g2	Female

The web-pages marked with rectangles in table 7.5 are missed by our model because some factors don't implemented and other limitation, we will discuss these missed webpages in the section "*Webpages Article Exceedance*" since they are common between males and females.

- *Job Experience Analysis:*

The demographic results regarding *years of experience in their job* is shown in table 7.6 and figure 7.10.

TABLE 7. 6: DEMOGRAPHIC RESULTS REGARDING EXPERIENCE.

Experience In Years	Less Than 2 Years	2 To 5 Years	6 To 10 Years	More Than 10 Years
Percentage (Group 1)	10%	30%	35%	25%
Percentage (Group 2)	10%	30%	20%	40%
Final Percentage	10%	30%	27.5%	32.5%

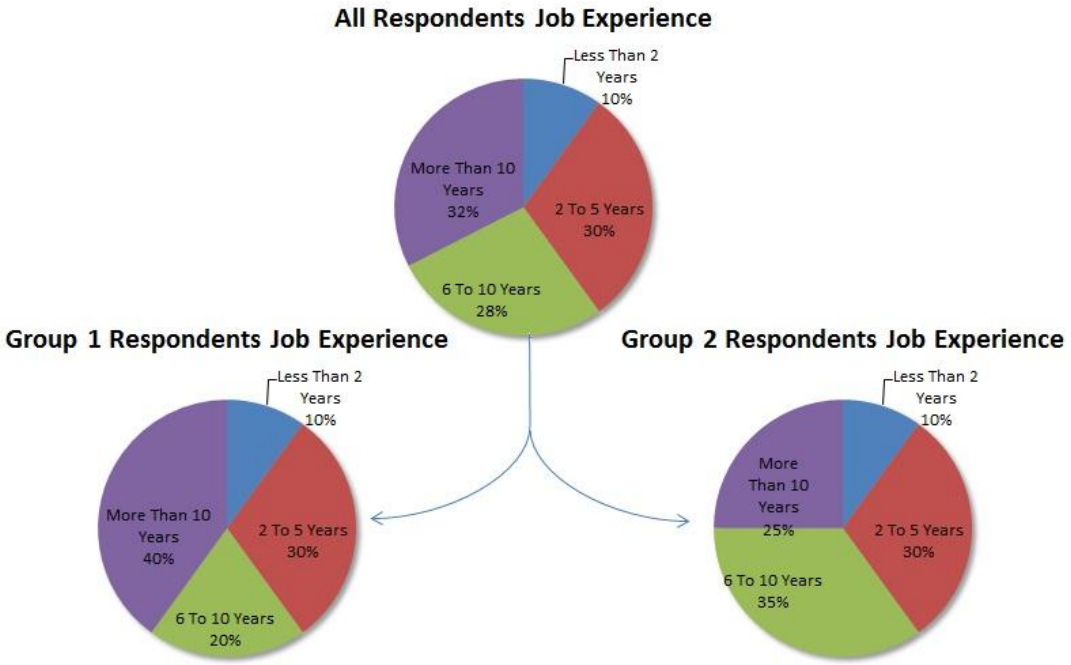


FIGURE 7. 10: RESPONDENTS JOB EXPERIENCE DISTRIBUTION.

We analyze the results according to job experience of 1-5 years' experience and more than 5, the results and their accuracy are shown in the figure 7.11(a) and 7.11(b) for group1 and the figures 7.11(c) and 7.11(d) for group2. They show no big difference between job experiences on each group. The missed webpages for each experience is shown in table 7.7, whereby webpages marked with rectangles are missed again for each group as happened in *gender analysis*.

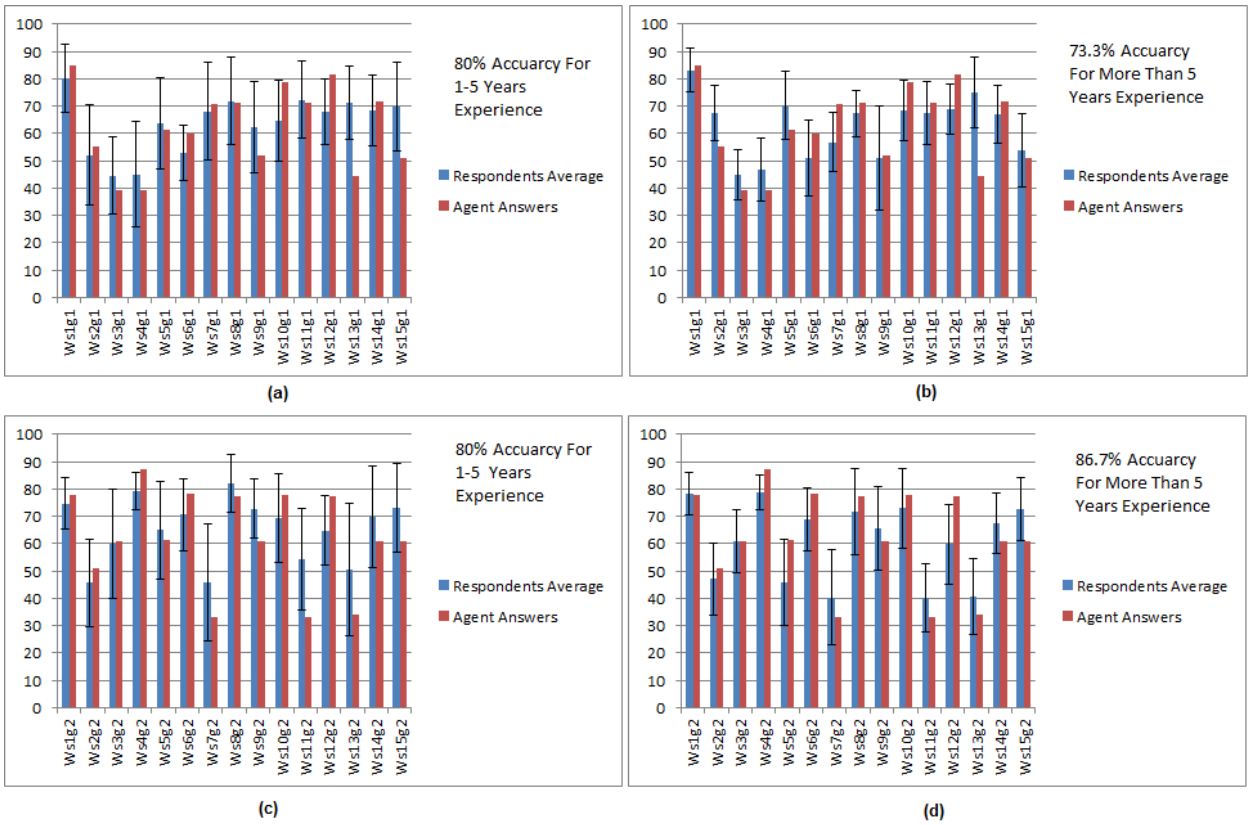


FIGURE 7.11 .: THE ACCURACY RESULTS FOR GROUP 1 (A): FOR 1-5 YEARS (B):MORE THAN 5 YEARS JOB EXPERIENCE. AND FOR GROUP 2 (C): FOR 1-5 YEARS (D):MORE THAN 5 YEARS JOB EXPERIENCE.

TABLE 7. 7: MISSED WEBPAGES SUMMARY ACCORDING TO JOB EXPERIENCE.

Group	Missed webpages	Job experience
1	Ws12g1, Ws13g1, Ws15g1	1-5 years
1	Ws2g1, Ws7g1, Ws12g1 and Ws13g1	More than 5 years
2	Ws4g2, Ws9g2, Ws11g2	1-5 years
2	Ws4g2, Ws12g2	More than 5 years

- *Academic Education Analysis:*

The academic educational distribution for respondents is shown in table 7.8 and figure 7.12, and we will study the effect of academic educational on our model accuracy. We study the group2 between "Engineering and Computer engineering/IT" and "Management/Financial /Accounting" respondents majors the results of accuracy is shown in figure 7.13(a) and (b). The figure shows no big difference in accuracy results for different academic types of respondents.

TABLE 7.8: DEMOGRAPHIC RESULTS REGARDING THE MAJORS.

Respondents Major	Group 1 Percentage	Group 2 Percentage	Final Percentage
Computer engineering/ IT	80%	20%	50%
Engineering	5%	20%	12.5%
Management/Financial/Accounting	10%	45%	27.5%
Other	5%	15%	10%

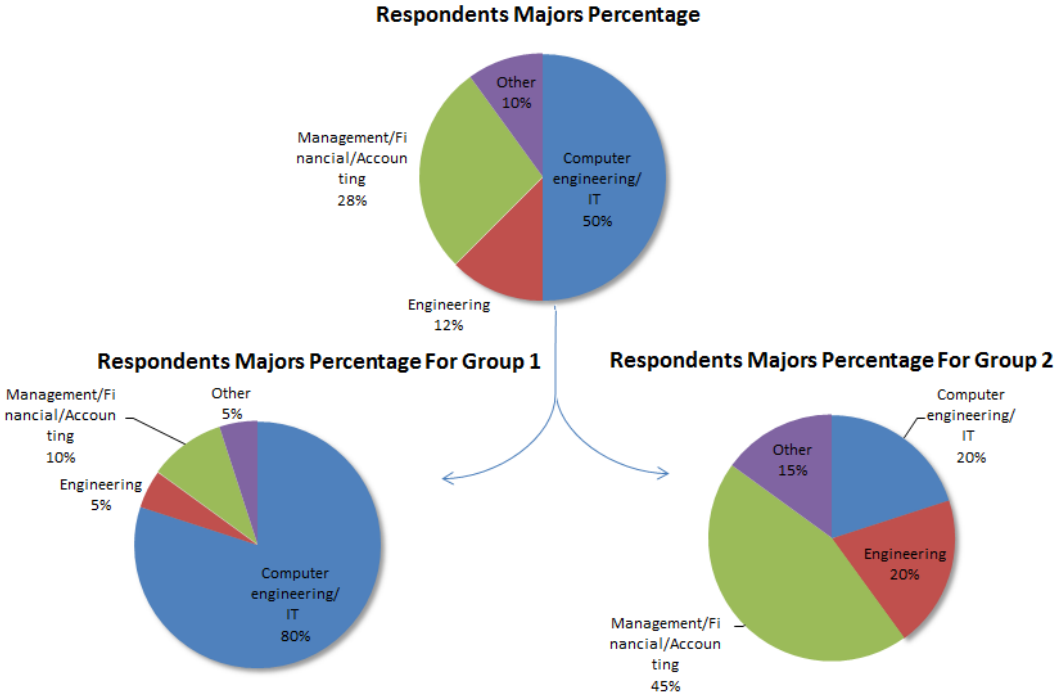


FIGURE 7. 12: RESPONDENTS MAJORS DISTRIBUTION.

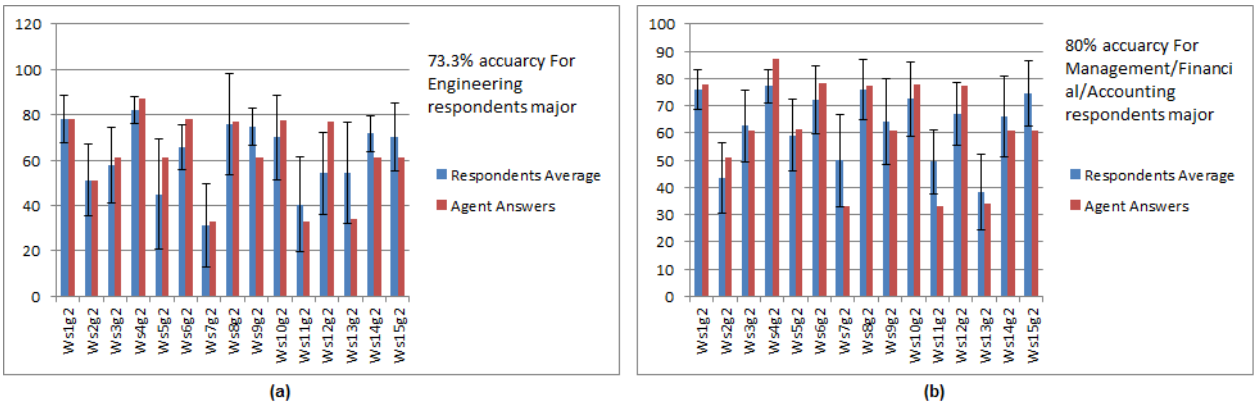


FIGURE 7. 13: (A) THE RESULTS AND ACCURACY OF ENGINEERING RESPONDENTS. (B) THE RESULTS AND ACCURACY OF MANAGEMENT/ FINANCIAL/ACCOUNTING RESPONDENTS

- *Webpages Article Exceedance:*

In two groups we notice that *Ws4g2, Ws12g1 and Ws13g1* are missed by CTA agent and exceed the confidence interval of average respondents' answers according to their gender, academic education even their majors, however this due to the lack of some factors implementation like referencing, neutralization… etc. that effect on accuracy of our results.

7.3.4 Factor Analysis Agents Collaboration Effects

- *MR and LR Factors Effects On HR:*

We used MR and LR factors give a supportive measure for HR factors. The slight effect of MR and LR factors on HR factors is shown in figure 7.14, table 7.9 for gruop1 and figure 7.15, table 7.10 for group2.

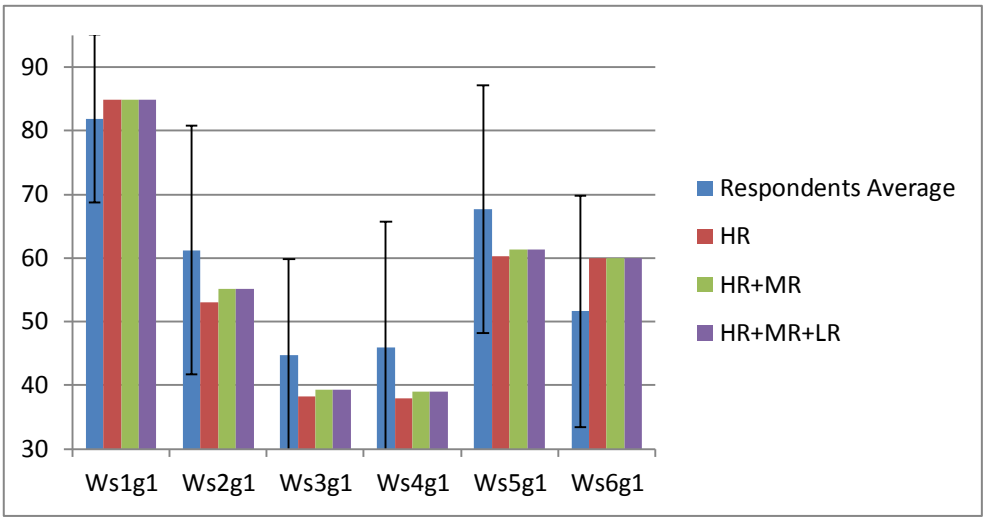


FIGURE 7. 14: EFFECT OF MR AND LR FACTORS ON HR FACTORS FOR GROUP1

TABLE 7.9: RESPONDENTS VS. CTA AGENT RESULTS FOR HR, MR AND LR (FIRST GROUP).

	Ws1g1	Ws2g1	Ws3g1	Ws4g1	Ws5g1	Ws6g1	Ws7g1	Ws8g1	Ws9g1	Ws10g1
CTA agent with HR	84.8	53	38.2	38	60.32	60	69.5	70.06	51.1	77.6
CTA agent with HR+MR	84.871	55.081	39.291	38.995	61.311	60.0011	70.566	71.105	52.081	78.661
CTA agent with HR+MR+LR	**84.872**	**55.0817**	**39.2911**	**38.9955**	**61.3115**	**60.0012**	**70.5667**	**71.1055**	**52.0811**	**78.662**
Respondents Average	**81.9**	**61.25**	**44.75**	**46**	**67.65**	**51.65**	**61.35**	**69.1**	**55.4**	**66.85**

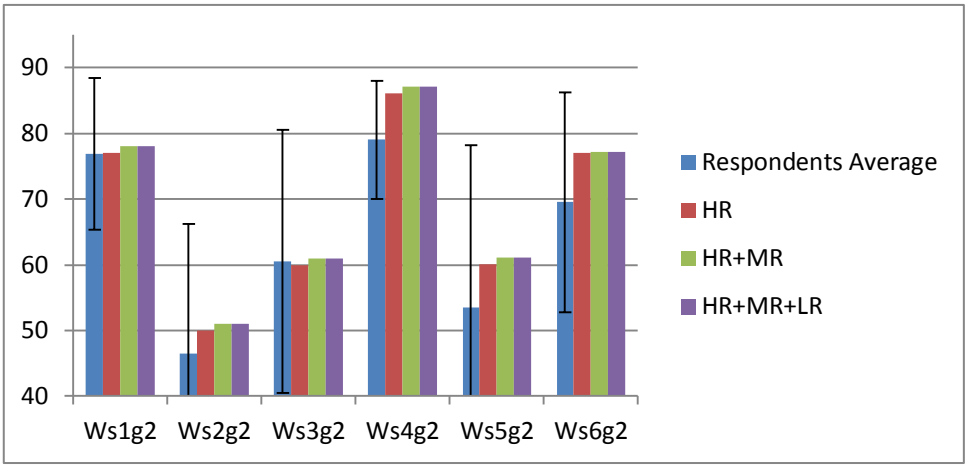


FIGURE 7. 15: EFFECT OF MR AND LR FACTORS ON HR FACTORS FOR GROUP2

TABLE 7. 10: RESPONDENTS VS. CTA AGENT RESULTS FOR HR, MR AND LR (SECOND GROUP).

	Ws1g1	Ws2g1	Ws3g1	Ws4g1	Ws5g1	Ws6g1	Ws7g1	Ws8g1	Ws9g1	Ws10g1
CTA agent with HR	77	50.021	60.0032	86	60.04	77.05	33.0003	77.0003	60.04001	77.6
CTA agent with HR+MR	77.972	51.092	60.9842	87.061	61.115	77.141	33.0983	77.0813	61.01701	77.691
CTA agent with HR+MR+LR	**77.973**	**51.0921**	**60.9847**	**87.062**	**61.0787**	**77.142**	**33.0993**	**77.0823**	**60.98071**	**77.692**
Respondents Average	**76.85**	**46.5**	**60.55**	**79**	**53.5**	**69.5**	**42.4**	**75.8**	**68.35**	**71.5**

- *Factor Analysis Agents Dependences:*

We analyze the relations between the content trust factors and how these factors could affect each other in order to improve trust judgments. This may generate some dependencies among factor analysis agents in our MAS making them collaborate and communicate away from CTA agents. In our work we study two cases of dependency between factor agents:

1- Popularity and Locality Relation:

As we mentioned in chapter3 the more *popular* resources are, the more they are trusted in their contents. However the user that reads information from a familiar resource, which is physically near to his place of residence, increase the trust judgment about these information. The popularity factor could be influenced by locality factor, whereby what is popular in your country could not necessarily be popular in other countries. The table 7.11 shows the possible value combination between these two factors; we compared the model judgment *(which popularity is dominant in our model)* with the judgment that generates the relation between these two factors:

TABLE 7. 11: THE EFFECT OF LOCALITY ON POPULARITY FACTOR

Popularity Value	Locality Value	Model Judgment	Relation Effect
High	High	High	High
Medium	High	Medium	*High*
Low	High	Low	*High*
High	Medium	High	High
Medium	Medium	Medium	Medium
Low	Medium	Low	Low
High	Low	High	High
Medium	Low	Medium	Medium
Low	Low	Low	Low

The special case that needs to be discussed here is when popularity agent factor value is "*low or medium*" while the locality factor agent value is "*high*." |In this case the locality value will be rationally dominant since the user who is close geographically to an information provides a high effect on trust judgment regardless of popularity value.

Handling this case in a webpage article will positively improve the trust judgment result approximately by 22.2%. Ws15g2 is an example of how the CTA agent judges this webpage as 74.42% that returns the webpage to its confidence interval [64.24, 81.26] and increasing the accuracy of gruop2 to 80%.

1- Specialization and Security Relation:

Another relation exists between specialization and security factors where this time shows the negative effects on trust judgment, the table 7.12 shows when this effect happens in a different combination of values. This combination is implemented in our model to improve its trust judgment results.

TABLE 7. 12: THE EFFECT OF SECURITY ON SPECIALIZATION FACTOR

Specialization Value	Security Value	Model Judgment	Relation Effect
Categorized	Minimal Risk	High	High
Categorized	Unverified	High	High
Categorized	Medium Risk	High	Medium
Categorized	High Risk	High	Low
Uncategorized	Minimal Risk	Low	Medium
Uncategorized	Unverified	Low	Medium
Uncategorized	Medium Risk	Low	Low
Uncategorized	High Risk	Low	Low

7.3.5 Case Study: Medical Webpages Content

We studied the model behavior on medical webpages by selecting 20 webpages that discuss "Breast Cancer Symptoms", webpages have all trust factors with different factor's values to generalize trust model on all factors cases. The survey was sent through Email to 30 doctors with different specialties while 20 doctors answered.

The figure 7.16 illustrates the results of trust model and how close to average doctors answers using confidence interval, to calculate the accuracy of model on medical webpages content, then:

$$Accuracy_{Medical} = \frac{Number\ of\ webpages\ within\ confidence\ interval}{Overall\ webpages\ count} \times 100\%$$

$$= \frac{17}{20} \times 100\% = 85\%$$

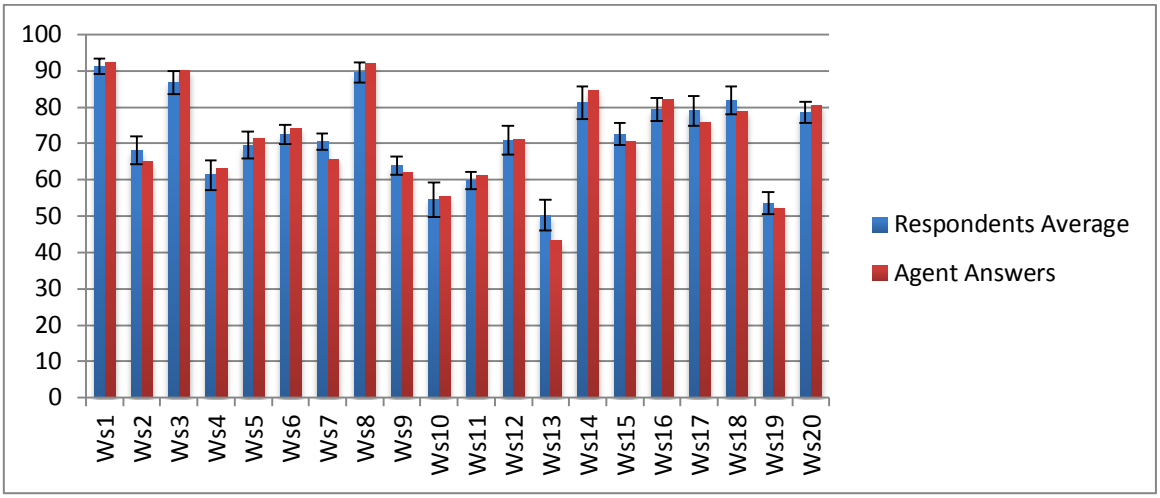


FIGURE 7. 16: CTA AGENT ANSWER COMPARED TO AVERAGE DOCTORS ANSWERS ON MEDICAL WEBPAGES.

The missed webpage articles *ws3*, *ws7* and *ws13* are not far away from their confidence interval boarders as appears precisely in table 7.13.

The confidence intervals become smaller when the medical topics is evaluated by doctors since their answers don't deviate significantly. The smaller intervals makes the trust model committed to provide more precision on its results as shown in figure 7.16, the precision returns to the nature of medical websites that belongs to web 1.0 sites where the trust model works more properly than web 2.0.

TABLE 7.13: RESPONDENTS RESULTS COMPARED TO CTA AGENT CALCULATION (MEDICAL ARTICLE).

	Ws1	Ws2	Ws3	Ws4	Ws5	Ws6	Ws7	Ws8	Ws9	Ws10	Ws11	Ws12	Ws13	Ws14	Ws15	Ws16	Ws17	Ws18	Ws19	Ws20
Respondent1	90	60	77	65	75	75	72	85	66	50	53	70	35	90	69	74	82	90	49	84
Respondent2	98	55	80	77	66	70	65	99	62	60	62	71	37	70	78	90	61	76	51	85
Respondent3	88	75	91	54	65	66	70	90	70	67	60	65	43	84	70	84	88	89	61	72
Respondent4	93	73	86	50	72	74	65	80	60	53	65	58	45	80	77	77	84	73	50	84
Respondent5	94	60	90	55	77	64	80	80	65	50	55	57	56	90	60	75	75	75	60	70
Respondent6	91	77	95	60	82	77	75	90	70	45	62	74	40	85	82	80	78	71	50	90
Respondent7	90	80	85	70	54	80	75	90	60	52	54	79	54	82	69	72	75	77	44	82
Respondent8	90	55	75	60	67	60	66	80	53	70	58	67	63	78	77	75	67	93	55	73
Respondent9	86	60	98	62	80	76	77	96	56	65	61	80	60	92	65	85	82	74	57	72
Respondent10	92	66	82	45	70	70	72	88	65	43	63	78	49	80	74	70	66	71	56	73
Respondent11	80	60	78	75	70	75	68	91	71	44	69	63	56	75	79	86	61	90	56	80
Respondent12	97	80	84	70	75	70	76	88	73	40	66	60	51	94	62	90	78	86	63	88
Respondent13	93	72	90	60	65	70	70	97	63	45	54	59	37	80	82	76	77	83	57	70
Respondent14	90	65	88	72	75	77	65	93	65	66	65	82	58	96	69	85	80	84	45	75
Respondent15	87	65	85	55	65	67	66	85	67	43	59	82	62	67	70	66	70	80	59	80
Respondent16	95	77	77	65	70	76	73	90	70	56	62	70	39	80	78	82	79	72	45	74
Respondent17	86	65	95	55	70	80	75	95	60	67	53	79	57	60	79	70	86	77	54	72
Respondent18	92	75	94	66	50	70	65	93	55	65	62	74	52	70	64	80	74	89	44	77
Respondent19	92	70	93	50	66	78	63	98	66	46	64	78	51	90	72	82	63	94	64	85
Respondent20	100	72	90	61	76	75	71	82	61	64	51	71	59	80	69	88	66	93	51	85
Respondents Average	**91.2**	**68.1**	**86.65**	**61.35**	**69.5**	**72.5**	**70.45**	**89.5**	**63.9**	**54.55**	**59.9**	**70.85**	**50.2**	**81.15**	**72.25**	**79.35**	**79**	**81.85**	**53.55**	**78.55**
Agent Answers	**92.1720**	**65.0728**	**90.0431**	**63.0632**	**71.3115**	**74.0743**	**65.5667**	**92.0211**	**62.0621**	**55.4921**	**61.0421**	**71.1852**	**43.2863**	**84.632**	**70.521**	**82.151**	**75.754**	**78.8**	**52.0953**	**80.4633**
Absolute Difference	0.9720	3.0272	3.3931	1.7132	1.8115	1.5743	4.8833	2.5211	1.8379	0.9421	1.1421	0.3352	6.9137	3.482	1.729	2.801	3.246	3.05	1.4547	1.9133
Standard Deviation σ	4.59519	8.05833	6.85392	8.73152	7.86397	5.43381	4.91480	6.03062	5.55261	10.0392	5.07729	8.31786	9.05887	9.49390	6.58447	7.02083	8.32497	8.13876	6.28678	6.46020
Confidence	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%
Critical t	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302	2.09302
Standard Error	1.02751	1.80189	1.53258	1.95242	1.75843	1.21503	1.09898	1.34848	1.24160	2.24484	1.13531	1.85993	2.02562	2.12290	1.47233	1.56990	1.86152	1.81988	1.40576	1.44454
Lower limit	89.0493	64.3285	83.4422	57.2635	65.8195	69.9568	68.1498	86.6775	61.3012	49.8514	57.5237	66.9571	45.9603	76.7067	69.1683	76.0641	75.1037	78.0409	50.6076	75.5265
Upper limit	93.3506	71.8714	89.8577	65.4364	73.1804	75.0431	72.7501	92.3224	66.4987	59.2485	62.2762	74.7428	54.4396	85.5932	75.3316	82.6358	82.8962	85.6590	56.4923	81.5734

: Missed webpages in our results

7.4 Results

1. Identify human perception factors of webpage content trust using empirical experiment that having collected trust factors implicitly to identify the order of factors importance from respondents' point of view.
2. Classification of factors to High, Medium and Low Rational factors based on their counts and calculated weights, for any factors Let f indicate a factor, $(f_1, f_2, f_3, \ldots f_n)$ are a set of $\boldsymbol{n}$ factors that exist in our trust model M, w_f is factor weight. f_i^j represents number of times a factor $\boldsymbol{i}$ ranked as $\boldsymbol{j}$, $j \in \{1, 2, \ldots K\}$, $i \in \{1, 2, \ldots n\}$ where K is the upper limit of ranking. R^j is partial weight for all factors ranked as $\boldsymbol{j}$. So, for any factor *i*:

$$R^j = \frac{\sum_{i=1}^{n} f_i^j}{\sum_{m=1}^{K} \sum_{i=1}^{n} f_i^m}$$

for each $j \in \{1, 2, \ldots K\}$

And its weight is:

$$w_{f_i} = f_i^1 R^1 + f_i^2 R^2 + \cdots + f_i^K R^K$$

The above formula can be abbreviated into the following formulas:

$$w_{f_i} = \sum_{j=1}^{K} f_i^j R^j$$

where $\sum_{j=1}^{K} R^j = 1$, for each factor $n \in \{1,2 \ldots N\}$

3. Developing a trust model based on factors classification that CTA agent rely on to calculate the trust value with factor analysis Agent Corporation.
4. The factor analysis agents provides precise representation *values* (v_f) about each article in our multi-agent system based on its knowledge that generated and updated using service, tools or a new extraction algorithms.
5. The multi-agent system consists of CTA agent and factor analysis agents that communicate with each other using either know-type messages or future-type messages depending on factor analysis agent classification (HR, MR and LR).
6. The validation process for trust model appear as rational results for all different 253 combination of factor analysis agent values $v_{f1}, v_{f2}, \ldots v_{fn} \epsilon [0, 9]$ depending on taking

three levels of factor values *(i.e., instead of taking all factors values from [0-9], we take low, medium and high)*. Here n=3 and k=21 then:

$$\binom{n+k-1}{k} = \binom{3+21-1}{21} = \binom{23}{21} = 253 \text{ Rational Possible Results}$$

7. From the evaluation experiment, the model of trust that exists in CTA agent provides a plausible vector of values μ_0 for the population mean vector μ for each webpage articles that have different trust factors types and levels. This verified by:

- ❖ Using the confidence intervals to check the percentage of how many webpages fall within their own confidence intervals. The results were 80% for group1 and 73.3% for group2.
- ❖ We used confidence intervals to determine which webpages have a bad judgmental value from CTA agent if they exceed it.
- ❖ The hypothesis H_0: $\mu = \mu_0$ for the p-dimensional respondents mean vector μ at level α=0.05 is ***rejected*** for group1 and ***accepted*** for group2.
- ❖ We ***accept*** the hypothesis H_0: $\mu = \mu_0$ for the p-dimensional respondents mean vector μ at level α=0.05 for group#1 after we exclude the webpages that exceed their confidence interval.

8. Using MR and LR factors values in our model provides a slightly supportive measure for HR factor values. However studying the factor analysis agents dependence's improve the trust judgment through positive or negative effects generate from the relations between these factor agents.

9. We analyze the relations between the content trust factors and how these factors could affect on each other in order to improve trust judgments. (*Popularity and Locality Relation, Specialization and Security Relation).*

Chapter Eight

Conclusions and Future Work

Chapter 8: Conclusion and Future Work

8.1 Conclusions

The trust of web information resource is considered one of the top hot research topics currently, so there is a strong need for multi-agent system that provides users over Internet a measurable value for some information or articles in websites.

- ❖ In this research, we introduce a new social driven trust model for information sources based on multi-agent system.
- ❖ We identify content trust factors using empirical experiment that identified these factors directly from respondents and showed how they are ordered according to their importance from respondents' point of view.
- ❖ We proposed a new mathematical formula to calculate factor weights to become prioritized, filtered and classifying them into three bands:
    - o High Rational (HR): where each factor belongs to this group has a critical effect on the system.
    - o Medium Rational (MR): where their values play supportive role for HR factors values.
    - o Low Rational (LR): where their values play supportive role for MR factors values.

Where their values in our model provide a slightly supportive measure for HR factors values.

- ❖ We introduce a new content trust algorithm representing the heart of our CTA agent, based on three bands.
- ❖ We introduce multi-agent system consisting of CTA agent and factor analysis agents that communicate with each other using either know-type messages or future-type messages.
- ❖ We evaluate the outputs of the trust model of our MAS that generate a trust value in a percentage and shows how close the trust value (accuracy) to human perception (from expert users). One method of evaluating the accuracy of output MAS is attained by comparing it with expert opinions (to keep following human perception especially expert users). So, we designed an empirical experiment based on expert population and prove that the model of trust existing in CTA agent provides a plausible vector of values μ0 for the

expert population mean vector μ for each webpage articles that have different trust factors types and levels. This verified by:

- Using the confidence intervals to check the percent how many webpages are fall within their own confidence intervals and it was 80% for group1 and 73.3% for group2.
- We used confidence interval to determine which webpages have a bad judgmental value from CTA agent if they exceed it.
- The hypothesis H_0: $\mu = \mu_0$ for the p-dimensional respondents mean vector μ at level α=0.05 is ***rejected*** for group1 and ***accepted*** for group2.
- We ***accept*** the hypothesis H_0: $\mu = \mu_0$ for the p-dimensional respondents mean vector μ at level α=0.05 for group#1 after we exclude the webpages that exceed their confidence interval.

8.2 Limitation

In our work there exists a set of limitation that must be mentioned. They and are as follows:

- The human perception factors need more and periodic investigation to be more reliable and comprehensive i.e. takes experts opinions, specific domain and web 2.0 compatibility.
- The factor analysis agents answer the CTA agent based on accumulated knowledge for a specific set of URLs since the process of calculation of a new URL request required very long calculation time. If their exists sufficient hardware, the calculation time of factor agents could be reduced form hours to minutes by applying some distribution systems algorithms and techniques.
- Appearance agent based on 5 expert opinions in its calculation that takes a very long calculation time; this is increase as the number of experts opinions increased. So it is also need a special hardware in order to reduce its calculation time.
- Some factors are hard to be captured or controlled on web pages since they need some standardization or new methodologies to be implemented like: Referencing, Seriously …etc.

8.3 Future Work

We propose an agent based trust model to determine the trust of content of web resources accuracy. This research has a lot of future works that could be done in order to either improve the model precision or develop on it. So, in the future work:

- A new human perception trust factors could appear or their rank could be changed with the passage of time. We can handle this case by periodic empirical experiments to control factors rank changes and make a comprehensive investigation for a new factors.
- The human perception factors could be merged with experts' opinions by making them ranking the factors. This method appears the true effect of some hidden factors which had not been notified by general humans perception i.e. security, repeating, rating… etc.
- Classifying the information in webpages based on its type or domain, will improves the judgment on content trust. Especially, if it is merged with a method that determine how much the information is important to be trusted, for example if a joke information exist in Facebook does not have the same trust importance if it was scientific.
- We can deeply analyze the relations between the content trust factors and how they could affect each other in order to improve trust judgments. This may generate some dependencies between factor analysis agents in our MAS making them collaborate and communicate a way from CTA agent.
- Classifying factor analysis agents into explicit and implicit factors helps to improve the trust judgment. The explicit factor is the factors that can be easily captured by a user when he reads an article or information on webpage like: appearance, specialization, …etc. while implicit factor cannot be easily captured by users like: security, authority, An intense study of the effect of these two types of factors classification improves the content trust value.
- We can develop a dynamic method to monitor the stability of respondent answers as the number of respondents' increases. This includes the ability to decide if the model could *accept or reject* someone's answers about an article based on studying the stability status in accept or reject cases.
- A dynamic classification method can provide us with wrong answers since it could be biased (it could be extended to accept new factors), can have revised factors classification by online questionnaires that have an ability to resist biasing. In our experiment we can control the bias because we control the environment.

- We can develop societies of agents that could contains other CTA agents, factor analysis agent and agents moving across the web, seeking new trust factors and then making them corporate with each other in better time and precision.
- In addition to human perception factors, we can deploy the information reported from *computer perception* to improve the accuracy of work like Google search engine as one of the agents in MAS. Another way of increasing the level of trust is attained by considering and calculating the *reputation* variable where it does not belong to human perception since it depends on previous experience.
- We can study how much each factor analysis agent is affected during time change (depending on absolute or relative time) and how could it affect this consideration on overall trust value.

References

[1] Xue, W., & Fan, Z. (2008). A New Trust Model Based on Social Characteristic and Reputation Mechanism for the Semantic Web. First International Workshop on Knowledge Discovery and Data Mining WKDD 2008, pp. 414-417.

[2] (ASF), A. S. (2012, October 25). DigestUtils. Retrieved October 26, 2012, from grepcode: http://grepcode.com/file/repo1.maven.org/maven2/commons-codec/commons-codec/1.5/org/apache/commons/codec/digest/DigestUtils.java?av=f

[3] Customer URL Ticketing System. (2011). Retrieved September 5, 2012, from McAfee TrustedSource™ : https://www.trustedsource.org/en/feedback/url?sid=B62C051B2979FF15A9EC4E0BC148C618

[4] XML Tutorial. (2011). Retrieved October 5, 2011, from w3schools: http://www.w3schools.com/xml/default.asp

[5] Agent-based model. (2013, January 3). Retrieved January 5, 2013, from wikipedia: http://en.wikipedia.org/wiki/Agent-based_model

[6] Abdul-Rahman, A., & Hailes, S. (1997). A Distributed Trust Model. Proceedings of the 1997 workshopon New security paradigms, pp. 48-60.

[7] Alexa. (2012). Top Sites. Retrieved September 2, 2012, from Alexa The Web Information Company: http://www.alexa.com/topsites

[8] Bernard , M. (2003, March 31). Criteria for optimal web design (designing for usability). Retrieved September 5, 2012, from http://www.optimalweb.org

[9] Castle, C., & Crooks, A. (2006). Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations. (CASA Working Papers). Centre for Advanced Spatial Analysis (UCL).

[10] Chang, C.-H., Kayed, M., Girgis, M. R., & Shaalan, K. (2006). A Survey of Web Information Extraction Systems. IEEE Transactions on Knowledge and Data Engineering, 18(10), pp. 1411-1428.

[11] Dondio, P., & Barrett, S. (2007, March 17). Computational Trust in Web Content Quality: A Comparative Evalutation on the Wikipedia Project. Informatica, 31, pp. 151-160.

[12] Ferrara, E., Fiumara, G., & Baumgartner , R. (2010, June). Web Data Extraction , Applications and Techniques : A Survey. ACM Computing Surveys, V, pp. 1-20.

[13] Gil , Y., & Artz, D. (2007). Towards Content Trust of Web Resources. Web Semantics: Science, Services and Agents on the World Widei Web, 5(4), pp. 227-239.

[14] Huynh, T. D., Jennings, N., & Shadbolt , N. (2004). Developing an integrated trust and reputation model for open multi-agent systems. Proceedings of the 7th International Workshop on Trust in Agent Societies , pp. 65-74.

[15] Huynh, T., Jennings, N., & Shadbolt, N. (2006). An integrated trust and reputation model for open multi-agent systems. Journal of Autonomous Agent Multi-Agent System, 13(2), pp. 119-154.

[16] IBM. (2002, March 14). Aglets. Retrieved Novermber 15, 2012, from IBM Corporation Research: http://www.research.ibm.com/trl/aglets/

[17] Idzelis, M., & Galbraith, B. (2012). Jazzy The Java Open Source Spell Checker. Retrieved November 13, 2012, from Jazzy Team: http://jazzy.sourceforge.net/

[18] Lam, M., Gong, Z., & Muyeba , M. (2008). A Method for Web Information Extraction. Progress in WWW Research and Development, 4976, pp. 383-394.

[19] Lange, D., Oshima, M., Karjoth, G., & Kosaka, K. (1997, March 10). Aglets: Programming Mobile Agents in Java. Springer, pp. 253-266.

[20] LanguageTool. (2012, November 29). HTTP API Using LanguageTool from Other Applications. Retrieved November 29, 2012, from LanguageTool: http://www.languagetool.org/http-api/

[21] linkchecker. (n.d.). LinkChecker. Retrieved October 20, 2012, from GoogleCode: http://code.google.com/p/linkchecker/

[22] McAfee. (2010). Reference Guide McAfee TrustedSource Web Database category set 4. Retrieved September 4, 2012, from McAfee TrustedSource™: https://www.trustedsource.org/download/ts_wd_reference_guide.pdf

[23] Sabater, J., & Sierra, C. (2001). REGRET: A reputation model for gregarious societies. on Deception Fraud and Trust in Agent Societies, 73(1), pp. 475-482.

[24] Teacy, W., Patel, J., Jennings, N., & Luck , M. (2006). TRAVOS: Trust and Reputation in the Context of Inaccurate Information Sources. Autonomous Agents and MultiAgent Systems, 12(2), pp. 183-198.

[25] Wang, W., Zeng, G., & Tang, D. (2010). Using evidence based content trust model for spam detection. Expert Systems with Applications, 37(8), pp. 5599–5606.

[26] Wikipedia. (2012, September 5). Categorization. Retrieved September 11, 2012, from Wikipedia: http://en.wikipedia.org/wiki/Categorization

[27] Wikipedia. (2012, September 5). PageRank. Retrieved September 10, 2012, from Wikipedia: http://en.wikipedia.org/wiki/PageRank

[28] Wikipedia. (2012, September 13). Secure Hash Algorithm. Retrieved September 16, 2012, from Wikipedia: http://en.wikipedia.org/wiki/Secure_Hash_Algorithm

[29] Wikipedia. (2012, September 10). Web traffic. Retrieved September 12, 2012, from Wikipedia: http://en.wikipedia.org/wiki/Web_traffic

[30] Marsh, S. P. (1994, April). Formalising Trust as a Computational Concept. PhD Thesis.

[31] Wikipedia. (2012, November 11). Information source. Retrieved November 14, 2012, from Wikipedia: http://en.wikipedia.org/wiki/Information_source

[32] Castelfranchi, C., Falcone, R., & Giovanni, P. (2003). Trust in information sources as a source for trust: a fuzzy approach. AAMAS '03: Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (pp. 89–96). NewYork, NY, USA: ACM Press.

[33] Moody, D. (2002, March 8). IT University of Copenhagen. Retrieved March 15, 2013, from IT University of Copenhagen Website: http://www.itu.dk/~oladjones/semester%203/advanced%20it%20mgt%20and%20software%20engineering/project/materials/what%20is%20empirical%20research1.pdf

[34] Pynadath, D., & Tambe, M. (2002). Multiagent teamwork: Analyzing key teamwork theories and models. *Proceedings of the first International Joint Conference on Autonomous Agents and Multi-Agent Systems*, (pp. 873–880).

[35] Kraus, S. (2001). *Strategic Negotiation in Multi-Agent Environments.* Cambridge MA, USA: MIT Press.

[36] Jennings, N., Faratin, P., Lomuscio, A., Parsons, S., Sierra, C., & Wooldridge, M. (2001). Automated Negotiation: Prospects, Methods and Challenges. *International Journal of Group Decision and Negotiation, 10*(2), pp. 199–215.

[37] Piper, P. S. (2000, September). *Better Read That Again: Web Hoaxes and Misinformation*. Retrieved April 5, 2013, from Information Today, Inc.: http://www.infotoday.com/searcher/sep00/piper.htm

[38] Kortum, P., Edwards, C., & Richards-Kortum, R. (2008). The Impact of Inaccurate Internet Health Information in a Secondary School Learning Environment. *JOURNAL OF MEDICAL INTERNET RESEARCH, 10*(2).

[39] Estabrook, L., & Rainie, L. (2007, December 30). *Information searches that solve problems. How people use the internet, libraries, and government agencies when they need help.* Retrieved April 2, 2013, from Pew Internet & American Life Project and University of

Illinois Libraries Survey: http://www.pewinternet.org/~/media/Files/Reports/2007/Pew_UI_LibrariesReport.pdf.pdf

[40] Lissman, T. L., & Boehnlein, J. K. (2001 , August 1). *A Critical Review of Internet Information About Depression*. Retrieved April 1, 2013, from The American Journal of Psychiatry: http://journals.psychiatryonline.org/article.aspx?Volume=52&page=1046&journalID=18

[41] Rodriguez, L., Insfran, E., & Cernuzzi, L. (2011). Requirements Modeling for Multi-Agent Systems. In F. Alkhateeb, E. Al Maghayreh, & I. Abu Doush (Eds.), *Multi-Agent Systems - Modeling, Control, Programming, Simulations and Applications* (p. 3). InTech.

[42] Webb , P. (n.d.). *Google's advice for your website content*. Retrieved February 2, 2012, from Webcredible: http://www.webcredible.co.uk/user-friendly-resources/search-engine-optimisation/google-content.shtml

[43] Press, O. U. (n.d.). *Definition of trust in English*. Retrieved July 10, 2013, from Oxford dictionaries: http://oxforddictionaries.com/definition/english/trust

[44] Mui, L. (2002, December 20). Computational Models of Trust and Reputation:Agents, Evolutionary Games, and Social Networks. *PhD Dissertation*, p. 75.

[45] Wikipedia. (2013, Febuary 5). *Scientific modelling*. Retrieved February 25, 2013, from Wikipedia: http://en.wikipedia.org/wiki/Scientific_modelling

[46] hibu. (2013, March 31). *Customers want trustworthy websites*. Retrieved April 22, 2013, from hibu business: http://business.hibu.co.uk/knowledge/articles/customers-want-trustworthy-websites/

[47] Charlton, G. (2010, February 17). *E-commerce trustmarks: do they matter?* Retrieved April 21, 2013, from econsultancy: http://econsultancy.com/us/blog /5303-e-commerce-trustmarks-do-they-matter

[48] VanFossen, L. (2012, February 6). *Prove It: What Makes You Trust a Website?* Retrieved April 18, 2013, from wordpress: http://lorelle.wordpress.com/2012/02/06/ prove-it-what-makes-you-trust-a-website/

[49] Aftermarketer Club. (2012, May 25). *24 Trust Factors That Improve Website Conversion Rates*. Retrieved April 18, 2013, from Aftermarketer Club Website: http://aftermarketerclub.com/blog/2012/05/24-trust-factors-that-improve-website-conversion-rates/

Appendix A

Appendix A: CTA Agent Java Code

```
public class CTAAgent extends Aglet {
    static Map<String, String> checkSumDB = new HashMap<String, String>();
       File factorsfilew = new File("E:\\ FactorAgentsData.csv ");
        FileWriter fwriter = new FileWriter(factorsfilew);

        public void onCreation(Object init) {
            System.out.println("CTA agent is created!");
        }
        public void run() {
        String vf;
        String line;
        int sample[] ,HR[]=null, MR[]=null, LR[]=null;
        sample =new int[7];
        HR =new int[3];
        MR =new int[3];
        LR =new int[2];
        int i=0,m,P;
        double res=0;
        boolean repeated=false,c=true;

        File filew = new File("E:\\CTAagentDB.csv");
        FileWriter writer = new FileWriter(filew);

        File data_file = new File("E:\\FactorAgentsData.csv");
        String url,app,cat,sec,pop,auth,age,mor,loc;
        float Trust_value=(float) 0.0;
        int loc_value=0;
        int
cat_value=0,pop_value=0,sec_value=0,age_value=0,app_value=0,mor_value=0,auth_v
alue=0,Trust_rank=0;

        System.out.println("Reading From Other Factor Analysis Agents ....");
            BufferedReader reader1 = new BufferedReader(new
FileReader(data_file));

              while((line=reader1.readLine())!=null) {
                  url=line.substring(0, line.indexOf(",")).trim();
                  line=line.substring(line.indexOf(",")+1).trim();
                  app=line.substring(0, line.indexOf(",")).trim();
                  line=line.substring(line.indexOf(",")+1).trim();
                  cat=line.substring(0, line.indexOf(",")).trim();
                  line=line.substring(line.indexOf(",")+1).trim();
                  auth=line.substring(0, line.indexOf(",")).trim();
                  line=line.substring(line.indexOf(",")+1).trim();
                  pop=line.substring(0, line.indexOf(",")).trim();
                  line=line.substring(line.indexOf(",")+1).trim();
                  age=line.substring(0,line.indexOf(",")).trim();
                  line=line.substring(line.indexOf(",")+1).trim();
                  mor=line.substring(0, line.indexOf(",")).trim();
                  sec=line.substring(line.indexOf(",")+1).trim();
                  line=line.substring(line.indexOf(",")+1).trim();
                  loc=line.substring(line.indexOf(",")+1).trim();

System.out.println(url+"\n"+cat+"\n"+sec+"\n"+pop+"\n"+auth);
                  if(cat.equals("Categorized URL")){
                      if(sec.equals("Minimal Risk"))
                      {
```

```
                        sample[6]=10;
                        sample[1]=7;
                    }
                else if(sec.equals("High Risk"))
                {
                    ///if category was malicious
                    sample[6]=1;
                    sample[1]=1;

                }
                else if(sec.equals("Unverified"))
                {
                    sample[6]=7;
                    sample[1]=6;
                }
                else if(sec.equals("Medium Risk"))
                {
                    sample[6]=5;
                    sample[1]=5;
                }
                else System.out.println("Error in Security..");

                }
                else if(cat.equals("Uncategorized URL"))
                {
                     if(sec.equals("Minimal Risk"))
                    {
                        sample[6]=10;
                        sample[1]=4;
                    }
                else if(sec.equals("High Risk"))
                {
                    sample[6]=1;
                    sample[1]=1;
                }
                else if(sec.equals("Unverified"))
                {
                    sample[6]=7;
                    sample[1]=4;
                }
                else if(sec.equals("Medium Risk"))
                {
                    sample[6]=5;
                    sample[1]=2;
                }
                else System.out.println("Error in Security..");

                }
                else System.out.println("Error in Cat..");
                sample[3]=(int) ((1.0-((float)
Integer.parseInt(pop)/1818594))*10+0.5);
                sample[4]=Integer.parseInt(age);
                sample[2]=Integer.parseInt(auth)-1;
                sample[0]=Integer.parseInt(app)-1;
                sample[5]=Integer.parseInt(mor);
                sample[7]=Integer.parseInt(loc);

                i=0;
            for(int j=0; j<3;j++){
                HR[j]=sample[j];
            }
```

```
     Arrays.sort(HR);
    for(int k=HR.length-1;k>=0;k--){
        m=2;
            while((m!=k)&& (m>=0)){
              if(HR[m]==HR[k]){
                  repeated=true;
              }
            m--;
            }
        if(!repeated){
            res+=(HR[k]*(Math.pow(10, HR[k])));
        }
        else{
            ////looking for available power
            int j=maxValue(HR);
            if(j == 0)
                res+=HR[k];
            else
                for(;j>0;j--){
                     if(!contains(HR,j-1)){
                         res+=(HR[k]*(Math.pow(10, j-1)));
                         c = false;
                     }

                     if(c){
                         res+=HR[k];
                          c=true;
                     }
                     else{
                         c=true;
                         break;
                     }

                  }

        }
        repeated=false;

    }

   ////////////////////MR Code/////////////////
        for(int j=3; j<6;j++)
             MR[j-3]=sample[j];
          Arrays.sort(MR);
          P=maxValue(HR)-3;
          for(int l=MR.length-1;l>=0 ;l--){
            res+=(MR[l]*(Math.pow(10, P)));
            P--;
          }
//////////////////LR Code/////////////////
         for(int j=6; j<7;j++)
             LR[j-6]=sample[j];
          Arrays.sort(LR);
          P=maxValue(HR)-5;
          for(int l=LR.length-1;l>=0;l--){
            res+=(LR[l]*(Math.pow(10, P)));
            P--;
          }

 res/=(Math.pow(10, maxValue(HR)));
writer.write(Arrays.toString(sample)+","+res+"\n");
```

```
        System.out.println(Arrays.toString(sample)+","+res);
        res=0;
        }
              writer.flush();
        writer.close();



    }
   private static int maxValue(int[] chars) {
        int max = chars[0];
        for (int ktr = 0; ktr < chars.length; ktr++) {
                if (chars[ktr] > max) {
                        max = chars[ktr];
                }
        }
        return max;
}
    private static boolean contains (int[] chars , int k) {

        for (int ktr = 0; ktr < chars.length; ktr++) {
                if (chars[ktr] == k) {
                        return true;
                }
        }
        return false;
}

}
        }
        public boolean handleMessage(Message msg) {
            if (msg.sameKind("appearance_agent") {
                String res=msg.getArg("vf");
               Fwriter.write(res+",");
               return true; // i know this message...
            } else if (msg.sameKind("specialization_agent") {
               String res=msg.getArg("vf");
               Fwriter.write(res+",");
               return true; // i know this message...
            } else if (msg.sameKind("authority_agent") {
                String res=msg.getArg("vf");
               Fwriter.write(res+",");
               return true; // i know this message...
            } else return false;
            return true;
        }
        public void onDisposing() {
            System.out.println("bye!");
        }
    }
```

Appendix B

Appendix B: McAfee TrustedSource™ Permission Letter

Subject	**[#KZJ-199-44096]: Permission Letter Request**
From	McAfee GTI Customer Response Team <sites@mcafee.com>,
To	<rmohsen@ppu.edu>,
Reply-To	<sites@mcafee.com>,
Date	17.12.2012 10:50
Priority	Normal

Hello Mr. Mohsen,

Thank you for your request. Please understand that we cannot provide the requested service for such an amount of URLs. You can use the customer ticketing portal for submission with your registered account. The daily maximum limit for Trusted Source customer submission is 100 URLs. The average turnaround time of the URLs submitted through Trusted Source is 24 hours. You will also be able to track the status of your URL submissions.

Thank you.

Petra Dornhof
Web Analyst
Customer Response Team - EMEA

McAfee GmbH
Vattmannstr. 3
33100 Paderborn
Germany
www.mcafee.com

Firmensitz: Muenchen
Amtsgericht: AG Muenchen
Handelsregister: HRB 144340
Geschaeftsfuehrer: Emmet Russell, Keith Krzeminski, Douglas Rice
Bankverbindung: ABN-Amro Bank N.V. Konto 671 211 9006
UST-ID: DE168122444

Ticket History

Client Posted On: 15 December 2012 07:26 PM

Dear McAfee Trusted Source Team,

I want first of all to thanks all McAfee team for all efforts to provide secure life over internet, I wonder if you can provide me a permission letter to use the McAfee SmartFilter XL database<https://www.trustedsource.org/en/feedback/url> (For Categorization and Reputation) on approximately 100,000 URLs to use them in my Master research thesis.

Thankful for your cooperation,

Appendix C

Appendix C: Model Validation Code and Results

- **Validation Design:**

The validation here aims to check whether the system software meets specifications and fulfills its intended purpose. To perform that, we used an Excel spreadsheet containing most relevant possible results from Combinations of 21 factor values each with a value between [0-9]. The number of results from this combinations *(with repetition)* of n=10 values taken k=21 times will be as following:

$$\binom{n+k-1}{k} = \binom{10+21-1}{21} = \binom{30}{21} = 14{,}307{,}150 \text{ Possible Results}$$

Since we cannot validate this huge number of possible results, we chose 253 results depending on taking three levels of factor values *(i.e., instead of taking all factors values from [0-9], we take low, medium and high)*. Here n=3 and k=21 then:

$$\binom{n+k-1}{k} = \binom{3+21-1}{21} = \binom{23}{21} = 253 \text{ Possible Results}$$

- **Validation Results:**

Appendix C shows all the results of validation process for the trust model. Table C.1 is a sample of validation in Appendix C, showing that our model behaves well according to factor value. In general, we can notice that the result becomes "high" when any factors from *f1* to *f4* (*which represents critical factors area*) becomes "high", and also when the results act as "Medium" and "Low" when these factors have "Medium" and "Low" values. From *f5* to *f8* (*which represents medium factors area*) plays the role of supportive factors for factors [*f1-f4*]; we can notice that from the fourth and fifth rows when the factors [*f1-f4*] values are equals. Finally, the rest of factors [*f9-f21*] (*which represents low factors area*) are supportive factors too, but with a lesser effect than medium factors on final trust result as seen in first and second rows.

TABLE C. 1: A SAMPLE OF MODEL VALIDATION.

f1	*f2*	*f3*	*f4*	*f5*	*f6*	*f7*	*f8*	*f9*	*f10*	*f11*	*f12*	*f13*	*f14*	*f15*	*f16*	*f17*	*f18*	*f19*	*f20*	*f21*	*Trust Percent*
0	0	0	1	4	4	5	6	6	7	8	8	8	8	8	8	8	8	9	9	9	10.65539989%
0	0	0	1	4	4	5	6	6	7	9	9	9	9	9	9	9	9	9	9	9	10.6554%
0	1	1	3	3	3	6	6	7	8	8	8	8	9	9	9	9	9	9	9	9	31.70%
0	1	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	6	41.45%
0	1	1	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	7	41.55%
0	1	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	9	51.56%
0	1	1	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	9	61.67%
0	1	1	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	8	8	9	71.78%
0	1	1	8	8	8	8	8	8	8	8	8	8	8	8	8	8	9	9	9	9	81.89%

- **Validation Code:**

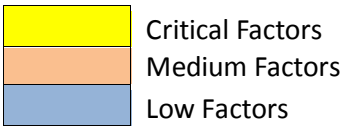
```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package agentvaildation;

import java.io.*;
import java.util.Arrays;


/**
 *
 * @author Rami
 */
public class AgentVaildation {

    /**
     * @param args the command line arguments
     */
    public static void main(String[] args) throws FileNotFoundException, IOException {
        // TODO code application logic here
        String vf;
        String line;
        int sample[] ,HR[]=null, MR[]=null, LR[]=null;
        sample =new int[21];
        HR =new int[4];
        MR =new int[4];
        LR =new int[13];
        int i=0,m,P;
        double res=0;
        boolean repeated=false,c=true;
        File data_file = new File("E:\\Valid\\2.csv");
        System.out.println("Reading Databases");
       ////////////////////
        File file = new File("E:\\Valid\\res12.csv");
                                // if file doesnt exists, then create it
            if (!file.exists()) {
            file.createNewFile();
            }

            FileWriter fw = new FileWriter(file.getAbsoluteFile());
            BufferedWriter bw = new BufferedWriter(fw);
       ////////////////////////////////
        BufferedReader reader1 = new BufferedReader(new FileReader(data_file));
        while((line=reader1.readLine())!=null) {
```

```
while(i<21){
    vf=line.substring(0, line.indexOf(",")).trim();
    line=line.substring(line.indexOf(",")+1).trim();

    sample[i]=Integer.parseInt(vf);

    System.out.println("Sample["+i+"]"+sample[i]+"Line:"+line);
    i++;
}
i=0;
for(int j=0; j<4;j++){
   HR[j]=sample[j];
}
Arrays.sort(HR);
for(int k=HR.length-1;k>=0;k--){
   m=3;
      while((m!=k)&& (m>=0)){
       if(HR[m]==HR[k]){
          repeated=true;
       }
      m--;
      }
   if(!repeated){
      res+=(HR[k]*(Math.pow(10, HR[k])));
   }
   else{
      ////looking for available power
      int j=maxValue(HR);
      if(j == 0)
         res+=HR[k];
      else
         for(;j>0;j--){
            if(!contains(HR,j-1)){
               res+=(HR[k]*(Math.pow(10, j-1)));
               c = false;
            }

            if(c){
               res+=HR[k];
                c=true;
            }
            else{
               c=true;
               break;
            }

         }

   }
```

```
            repeated=false;

          }
        ////////////////////MR Code/////////////////
             for(int j=4; j<8;j++)
                MR[j-4]=sample[j];
              Arrays.sort(MR);
              P=maxValue(HR)-2;
            //   if(P>0)
              for(int l=MR.length-1;l>=0 ;l--){
                res+=(MR[l]*(Math.pow(10, P)));
                P--;
              }
        ////////////////////LR Code/////////////////
             for(int j=8; j<21;j++)
                LR[j-8]=sample[j];
              Arrays.sort(LR);
              P=maxValue(HR)-5;
             // if(P>0)
              for(int l=LR.length-1;l>=0;l--){
                res+=(LR[l]*(Math.pow(10, P)));
                P--;
              }
       res/=(Math.pow(10, maxValue(HR)));
          bw.write(Arrays.toString(sample)+","+res+"\n");
          System.out.println(Arrays.toString(sample)+","+res);
       res=0;
       }
       bw.close();

    }
    private static int maxValue(int[] chars) {
          int max = chars[0];
          for (int ktr = 0; ktr < chars.length; ktr++) {
                    if (chars[ktr] > max) {
                              max = chars[ktr];
                    }
          }
          return max;
}
    private static boolean contains (int[] chars , int k) {

          for (int ktr = 0; ktr < chars.length; ktr++) {
                    if (chars[ktr] == k) {
                              return true;
                    }
          }
          return false;
}
}
```

- **Selected Validation Results:**

	Factor
f1	Referencing
f2	Appearance of Website
f3	Authority
f4	Specialization
f5	The Methodology of Writing
f6	Seriously
f7	Age and Update
f8	Popularity
f9	Multimedia Supporting
f10	User Expertise
f11	Proximity
f12	Locality
f13	Links Inconsistency
f14	Evaluating and Comments
f15	Neutralization
f16	Contact
f17	Constrains
f18	Feedback
f19	Security
f20	Number Of Visitors
f21	Repeating



f1	*f2*	*f3*	*f4*	*f5*	*f6*	*f7*	*f8*	*f9*	*f10*	*f11*	*f12*	*f13*	*f14*	*f15*	*f16*	*f17*	*f18*	*f19*	*f20*	*f21*	***Trust Percent***
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	5	0.10%
0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	0.11%
0	0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	2	4	0.20%
0	0	0	0	0	0	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0.21%
0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0.22%
0	0	0	0	0	0	0	3	3	3	3	3	3	3	3	3	3	3	3	3	5	0.30%
0	0	0	0	0	0	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0.31%
0	0	0	0	0	0	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0.32%
0	0	0	0	0	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0.33%
0	0	0	0	0	0	0	4	4	4	4	4	4	4	4	4	4	4	4	5	5	0.40%
0	0	0	0	0	0	1	4	4	4	4	4	4	4	4	4	4	4	4	4	8	0.41%

0	0	0	0	0	0	2	4	4	4	4	4	4	4	4	4	4	4	4	4	8	0.42%
0	0	0	0	0	0	3	4	4	4	4	4	4	4	4	4	4	4	4	4	8	0.43%
0	0	0	0	0	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0.44%
0	0	0	0	0	0	0	5	5	5	5	5	5	5	5	5	5	5	5	6	7	0.50%
0	0	0	0	0	0	1	5	5	5	5	5	5	5	5	5	5	5	5	5	9	0.51%
0	0	0	0	0	0	2	5	5	5	5	5	5	5	5	5	5	5	5	5	9	0.52%
0	0	0	0	0	0	3	5	5	5	5	5	5	5	5	5	5	5	5	5	9	0.53%
0	0	0	0	0	0	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	0.54%
0	0	0	0	0	0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	9	0.55%
0	0	0	0	0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	7	0.56%
0	0	0	0	0	0	0	6	6	6	6	6	6	6	6	6	6	6	6	6	8	0.60%
0	0	0	0	0	0	1	6	6	6	6	6	6	6	6	6	6	6	6	6	6	0.61%
0	0	0	0	0	0	2	6	6	6	6	6	6	6	6	6	6	6	6	6	6	0.62%
0	0	0	0	0	0	3	6	6	6	6	6	6	6	6	6	6	6	6	6	6	0.63%
0	0	0	0	0	0	4	6	6	6	6	6	6	6	6	6	6	6	6	7	7	0.64%
0	0	0	0	0	0	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	0.65%
0	0	0	0	0	0	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	0.66%
0	0	0	0	0	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	8	0.67%
0	0	0	0	0	0	0	7	7	7	7	7	7	7	7	7	7	7	7	7	9	0.70%
0	0	0	0	0	0	1	7	7	7	7	7	7	7	7	7	7	7	7	7	7	0.71%
0	0	0	0	0	0	2	7	7	7	7	7	7	7	7	7	7	7	7	7	7	0.72%
0	0	0	0	0	0	3	7	7	7	7	7	7	7	7	7	7	7	7	7	7	0.73%
0	0	0	0	0	0	4	7	7	7	7	7	7	7	7	7	7	7	7	8	9	0.74%
0	0	0	0	0	0	5	7	7	7	7	7	7	7	7	7	7	7	7	7	7	0.75%
0	0	0	0	0	0	6	7	7	7	7	7	7	7	7	7	7	7	7	7	7	0.76%
0	0	0	0	0	0	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	0.77%
0	0	0	0	0	5	7	7	7	7	7	7	7	7	7	7	7	7	7	7	9	0.78%
0	0	0	0	0	0	0	8	8	8	8	8	8	8	8	8	8	8	8	8	9	0.80%
0	0	0	0	0	0	1	8	8	8	8	8	8	8	9	9	9	9	9	9	9	0.81%
0	0	0	0	0	0	2	8	8	8	8	8	8	8	9	9	9	9	9	9	9	0.82%
0	0	0	0	0	0	3	8	8	8	8	8	8	8	9	9	9	9	9	9	9	0.83%
0	0	0	0	0	0	4	8	8	8	8	8	8	8	8	8	8	8	9	9	9	0.84%
0	0	0	0	0	0	5	8	8	8	8	8	8	8	9	9	9	9	9	9	9	0.85%
0	0	0	0	0	0	6	8	8	8	8	8	8	8	9	9	9	9	9	9	9	0.86%
0	0	0	0	0	0	7	8	8	8	8	8	8	8	9	9	9	9	9	9	9	0.87%
0	0	0	0	0	0	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	0.88%
0	0	0	0	0	5	8	8	8	8	8	8	8	8	8	8	8	8	8	9	9	0.89%
0	0	0	0	1	1	1	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0.91%
0	0	0	0	1	1	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0.92%
0	0	0	0	1	1	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0.93%
0	0	0	0	1	1	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0.95%
0	0	0	0	1	1	6	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0.96%
0	0	0	0	1	1	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0.97%
0	0	0	0	0	1	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0.98%
0	0	0	0	1	1	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0.99%

0	0	0	0	1	4	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	1.00%
0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	10.11%
0	0	0	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	10.21%
0	0	0	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	10.22%
0	0	0	1	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	10.31%
0	0	0	1	1	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	10.32%
0	0	0	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	10.33%
0	0	0	1	1	1	1	4	4	4	4	4	4	4	4	4	4	4	4	4	8	10.41%
0	0	0	1	1	1	2	4	4	4	4	4	4	4	4	4	4	4	4	4	8	10.42%
0	0	0	1	1	1	3	4	4	4	4	4	4	4	4	4	4	4	4	4	8	10.43%
0	0	0	1	1	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	10.44%
0	0	0	1	1	1	1	5	5	5	5	5	5	5	5	5	5	5	5	5	9	10.51%
0	0	0	1	1	1	2	5	5	5	5	5	5	5	5	5	5	5	5	5	9	10.52%
0	0	0	1	1	1	3	5	5	5	5	5	5	5	5	5	5	5	5	5	9	10.53%
0	0	0	1	1	1	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	10.54%
0	0	0	1	1	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	9	10.55%
0	0	0	1	1	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	9	10.56%
0	0	0	1	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	10.56%
0	0	0	1	1	1	1	6	6	6	6	6	6	6	6	6	6	6	6	6	6	10.61%
0	0	0	1	1	1	2	6	6	6	6	6	6	6	6	6	6	6	6	6	6	10.62%
0	0	0	1	1	1	3	6	6	6	6	6	6	6	6	6	6	6	6	6	6	10.63%
0	0	0	1	1	1	4	6	6	6	6	6	6	6	6	6	6	6	6	7	7	10.64%
0	0	0	1	1	1	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	10.65%
0	0	0	1	4	4	5	6	7	7	7	7	7	7	8	8	8	8	8	9	9	10.66%
0	0	0	1	1	1	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	10.66%
0	0	0	1	1	5	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	10.67%
0	0	0	1	1	1	1	7	7	7	7	7	7	7	7	7	7	7	7	7	7	10.71%
0	0	0	1	1	1	2	7	7	7	7	7	7	7	7	7	7	7	7	7	7	10.72%
0	0	0	1	1	1	3	7	7	7	7	7	7	7	7	7	7	7	7	7	7	10.73%
0	0	0	1	1	1	4	7	7	7	7	7	7	7	7	7	7	7	7	8	9	10.74%
0	0	0	1	1	1	5	7	7	7	7	7	7	7	7	7	7	7	7	7	7	10.75%
0	0	0	1	4	4	5	7	7	7	7	7	7	7	7	7	7	7	7	9	9	10.76%
0	0	0	1	1	1	6	7	7	7	7	7	7	7	7	7	7	7	7	7	7	10.76%
0	0	0	1	1	1	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	10.77%
0	0	0	1	1	5	7	7	7	7	7	7	7	7	7	7	7	7	7	8	9	10.78%
0	0	0	1	1	1	1	8	8	8	8	8	8	8	9	9	9	9	9	9	9	10.81%
0	0	0	1	1	1	2	8	8	8	8	8	8	8	9	9	9	9	9	9	9	10.82%
0	0	0	1	1	1	3	8	8	8	8	8	8	8	9	9	9	9	9	9	9	10.83%
0	0	0	1	1	1	4	8	8	8	8	8	8	8	8	8	8	8	9	9	9	10.84%
0	0	0	1	1	1	5	8	8	8	8	8	8	8	9	9	9	9	9	9	9	10.85%
0	0	0	1	4	4	5	8	8	8	8	8	8	8	8	8	8	9	9	9	9	10.86%
0	0	0	1	1	1	6	8	8	8	8	8	8	8	9	9	9	9	9	9	9	10.86%
0	0	0	1	1	1	7	8	8	8	8	8	8	8	9	9	9	9	9	9	9	10.87%
0	0	0	1	1	1	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	10.88%
0	0	0	1	1	5	8	8	8	8	8	8	8	8	8	8	8	9	9	9	9	10.89%

0	0	0	1	2	2	4	9	9	9	9	9	9	9	9	9	9	9	9	9	9	10.94%
0	0	0	1	2	4	4	9	9	9	9	9	9	9	9	9	9	9	9	9	9	10.95%
0	0	0	1	2	5	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	10.96%
0	0	0	1	2	5	6	9	9	9	9	9	9	9	9	9	9	9	9	9	9	10.97%
0	0	0	1	2	5	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	10.98%
0	0	0	1	3	4	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	10.99%
0	0	0	1	2	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	11.00%
0	0	0	1	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	11.00%
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	11.11%
0	0	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	3	11.21%
0	0	1	1	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	4	11.31%
0	0	1	1	1	1	1	4	4	4	4	4	4	4	4	4	4	4	4	4	9	11.41%
0	0	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5	5	5	6	6	11.51%
0	0	1	1	1	1	1	6	6	6	6	6	6	6	6	6	6	6	6	6	7	11.61%
0	0	1	1	1	1	1	7	7	7	7	7	7	7	7	7	7	7	7	7	8	11.71%
0	0	1	1	1	1	1	8	8	8	8	8	8	8	8	8	8	8	8	8	8	11.81%
0	0	1	1	1	1	5	8	8	8	8	8	8	8	8	8	8	8	8	8	8	11.85%
0	0	1	1	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	12.00%
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	12.11%
0	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	8	12.21%
0	1	1	1	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	9	12.31%
0	1	1	1	1	1	1	4	4	4	4	4	4	4	4	4	4	4	4	4	6	12.41%
0	1	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5	5	5	5	7	12.51%
0	1	1	1	1	1	1	6	6	6	6	6	6	6	6	6	6	6	6	7	9	12.61%
0	1	1	1	1	1	1	7	7	7	7	7	7	7	7	7	7	7	8	8	8	12.71%
0	1	1	1	1	1	1	8	8	8	8	8	8	8	8	8	9	9	9	9	9	12.81%
0	1	1	1	1	1	5	8	8	8	8	8	8	8	8	8	9	9	9	9	9	12.85%
0	1	1	1	1	1	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	12.98%
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	13.00%
0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	9	20.22%
0	0	0	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	4	4	20.32%
0	0	0	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	7	20.42%
0	0	0	2	2	2	2	5	5	5	5	5	5	5	5	5	5	5	5	5	8	20.52%
0	0	0	2	2	2	2	6	6	6	6	6	6	6	6	6	6	6	6	8	8	20.62%
0	0	0	2	2	2	2	7	7	7	7	7	7	7	7	7	7	7	8	8	9	20.72%
0	0	0	2	2	2	2	8	8	8	8	8	8	8	9	9	9	9	9	9	9	20.82%
0	0	0	2	2	2	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	20.92%
0	0	0	2	2	2	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	20.95%
0	0	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	6	21.22%
0	0	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	7	21.32%
0	0	1	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	21.42%
0	0	1	2	2	2	2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	21.52%
0	0	1	2	2	2	2	6	6	6	6	6	6	6	6	6	6	6	6	7	7	21.62%
0	0	1	2	2	2	2	7	7	7	7	7	7	7	7	7	7	7	7	8	9	21.72%
0	0	1	2	2	2	2	8	8	8	8	8	8	8	8	8	8	8	9	9	9	21.82%

0	0	1	2	2	2	5	8	8	8	8	8	8	8	8	8	8	8	9	9	9	21.85%
0	0	1	2	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	22.00%
0	1	1	2	2	2	5	8	8	8	8	8	8	8	8	8	8	8	8	8	8	22.05%
0	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	22.22%
0	0	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	22.32%
0	0	2	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	8	22.42%
0	0	2	2	2	2	2	5	5	5	5	5	5	5	5	5	5	5	5	5	9	22.52%
0	0	2	2	2	2	2	6	6	6	6	6	6	6	6	6	6	6	6	6	6	22.62%
0	0	2	2	2	2	2	7	7	7	7	7	7	7	7	7	7	7	7	7	7	22.72%
0	0	2	2	2	2	2	8	8	8	8	8	8	8	9	9	9	9	9	9	9	22.82%
0	0	2	2	2	2	5	8	8	8	8	8	8	8	9	9	9	9	9	9	9	22.85%
0	0	2	2	3	5	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	22.96%
0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	24.00%
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	26.00%
0	0	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	30.33%
0	0	0	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	7	30.43%
0	0	0	3	3	3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	8	30.53%
0	0	0	3	3	3	3	6	6	6	6	6	6	6	6	6	6	6	6	8	8	30.63%
0	0	0	3	3	3	3	7	7	7	7	7	7	7	7	7	7	7	8	8	9	30.73%
0	0	0	3	3	3	3	8	8	8	8	8	8	8	9	9	9	9	9	9	9	30.83%
0	0	0	3	3	3	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	30.93%
0	0	0	3	3	3	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	30.95%
0	0	1	3	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	31.10%
0	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	31.43%
0	1	1	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	9	31.53%
0	1	1	3	3	3	3	5	5	5	5	5	5	5	5	5	5	5	5	6	6	31.63%
0	1	1	3	3	3	3	6	6	6	6	6	6	6	6	6	6	6	6	6	7	31.73%
0	1	1	3	3	3	3	7	7	7	7	7	7	7	7	7	7	7	7	7	8	31.83%
0	1	1	3	3	3	3	8	8	8	8	8	8	8	8	8	8	8	8	8	8	31.93%
0	1	1	3	3	3	5	8	8	8	8	8	8	8	8	8	8	8	8	8	8	31.95%
0	0	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	32.33%
0	0	2	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	8	32.43%
0	0	2	3	3	3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	9	32.53%
0	0	2	3	3	3	3	6	6	6	6	6	6	6	6	6	6	6	6	6	6	32.63%
0	0	2	3	3	3	3	7	7	7	7	7	7	7	7	7	7	7	7	7	7	32.73%
0	0	2	3	3	3	3	8	8	8	8	8	8	8	9	9	9	9	9	9	9	32.83%
0	0	2	3	3	3	5	8	8	8	8	8	8	8	9	9	9	9	9	9	9	32.85%
0	0	2	3	4	5	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	32.96%
0	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	7	33.33%
0	0	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	33.43%
0	1	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	5	33.50%
0	0	3	3	3	3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	33.53%
0	0	3	3	3	3	3	6	6	6	6	6	6	6	6	6	6	6	6	7	7	33.63%
0	0	3	3	3	3	3	7	7	7	7	7	7	7	7	7	7	7	7	8	9	33.73%
0	0	3	3	3	3	3	8	8	8	8	8	8	8	8	8	8	8	9	9	9	33.83%

0	0	3	3	3	3	5	8	8	8	8	8	8	8	8	8	8	8	9	9	9	33.85%
0	0	3	3	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	34.00%
1	1	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	34.50%
0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	36.30%
0	3	3	3	3	3	3	5	5	5	5	5	5	5	5	5	5	5	5	6	6	36.50%
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	7	39.30%
0	0	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	8	40.45%
0	0	0	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	9	40.55%
0	0	0	4	4	4	4	6	6	6	6	6	6	6	6	6	6	6	6	6	6	40.65%
0	0	0	4	4	4	4	7	7	7	7	7	7	7	7	7	7	7	7	7	7	40.75%
0	0	0	4	4	4	4	8	8	8	8	8	8	8	9	9	9	9	9	9	9	40.85%
0	0	0	4	4	4	5	8	8	8	8	8	8	8	8	8	8	8	9	9	9	40.86%
0	0	0	4	4	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	41.00%
0	0	2	4	4	4	5	8	8	8	8	8	8	8	8	8	8	9	9	9	9	41.06%
0	0	2	4	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	41.20%
0	0	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	43.44%
0	0	3	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	6	43.55%
0	0	3	4	4	4	4	6	6	6	6	6	6	6	6	6	6	6	6	7	8	43.65%
0	0	3	4	4	4	4	7	7	7	7	7	7	7	7	7	7	7	7	9	9	43.75%
0	0	3	4	4	4	4	8	8	8	8	8	8	8	8	8	8	9	9	9	9	43.85%
0	0	3	4	4	4	5	8	8	8	8	8	8	8	8	8	8	8	8	8	8	43.86%
0	0	3	4	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	44.00%
0	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	44.44%
0	0	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	44.54%
0	1	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	6	44.55%
0	0	4	4	4	4	4	6	6	6	6	6	6	6	6	6	6	6	6	7	7	44.65%
0	0	4	4	4	4	4	7	7	7	7	7	7	7	7	7	7	7	7	8	9	44.75%
0	0	4	4	4	4	4	8	8	8	8	8	8	8	8	8	8	8	9	9	9	44.85%
0	0	4	4	4	4	5	8	8	8	8	8	8	8	9	9	9	9	9	9	9	44.86%
0	0	4	4	5	5	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	44.96%
1	1	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	45.55%
2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	6	46.64%
0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	9	48.44%
0	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	6	6	48.54%
0	0	0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	7	50.56%
0	0	0	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	8	50.66%
0	0	0	5	5	5	5	7	7	7	7	7	7	7	7	7	7	7	7	7	9	50.76%
0	0	0	5	5	5	5	8	8	8	8	8	8	8	8	8	8	8	8	8	9	50.86%
0	0	0	5	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	51.00%
0	0	3	5	5	5	5	7	7	7	7	7	7	7	7	7	7	7	8	8	9	51.06%
0	0	3	5	5	5	5	8	8	8	8	8	8	8	8	9	9	9	9	9	9	51.16%
0	0	3	5	5	5	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	51.26%
0	0	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	54.56%
0	0	4	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	7	9	54.66%
0	0	4	5	5	5	5	7	7	7	7	7	7	7	7	7	7	7	8	8	8	54.76%

0	0	4	5	5	5	5	8	8	8	8	8	8	8	8	8	9	9	9	9	9	54.86%
0	0	4	5	5	5	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	54.99%
0	0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	8	55.56%
0	0	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	9	55.66%
0	0	5	5	5	5	5	7	7	7	7	7	7	7	7	7	7	7	7	8	8	55.76%
0	0	5	5	5	5	5	8	8	8	8	8	8	8	8	8	8	8	8	9	9	55.86%
0	0	5	5	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	56.00%
1	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	8	56.56%
2	2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	57.58%
3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	8	58.86%
0	0	0	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	8	60.67%
0	0	0	6	6	6	6	7	7	7	7	7	7	7	7	7	7	7	7	9	9	60.77%
0	0	0	6	6	6	6	8	8	8	8	8	8	8	8	8	8	9	9	9	9	60.87%
0	0	1	6	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	61.00%
0	0	4	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	61.07%
0	0	4	6	6	6	6	7	7	7	7	7	7	7	7	7	7	7	7	7	8	61.17%
0	0	4	6	6	6	6	8	8	8	8	8	8	8	8	8	8	8	8	8	8	61.27%
0	0	4	6	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	61.40%
0	0	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	9	65.67%
0	0	5	6	6	6	6	7	7	7	7	7	7	7	7	7	7	7	8	8	8	65.77%
0	0	5	6	6	6	6	8	8	8	8	8	8	8	8	8	9	9	9	9	9	65.87%
0	0	5	6	6	6	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	65.99%
0	0	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	66.67%
0	0	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7	7	7	8	9	66.77%
0	0	6	6	6	6	6	8	8	8	8	8	8	8	8	8	8	8	9	9	9	66.87%
1	1	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	67.67%
2	2	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	9	68.67%
0	0	0	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	8	70.78%
0	0	0	7	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8	8	8	70.88%
0	0	0	7	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	71.00%
0	0	5	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	9	71.28%
0	0	5	7	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8	8	9	71.38%
0	0	5	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	71.50%
0	0	6	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	76.78%
0	0	6	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9	9	9	76.88%
0	0	6	7	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	77.00%
0	0	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	9	77.78%
0	0	7	7	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8	8	9	77.88%
0	0	7	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	78.00%
0	5	7	7	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	78.50%
0	0	0	8	8	8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	80.89%
0	0	0	8	8	8	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	80.99%
0	0	6	8	8	8	8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	81.49%
0	0	7	8	8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9	87.89%

0	0	7	8	8	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	88.00%
0	0	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	9	9	9	9	88.89%
0	6	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	89.59%
0	0	4	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	91.00%
3	3	4	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	94.00%
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	100.00%
1	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	100.00%

Appendix D

Appendix D: Evaluation Questionnaire

Group #1 Questionnaire

Edit this form

استبيان حول الثقة بمحتوى المعلومات في المواقع الالكترونية

عزيزي القارئ/ عزيزتي القارئة:
يقوم الباحث بإجراء دراسة لمعرفة العوامل الأساسية المؤثرة في الثقة بمحتوى المعلومات الموجودة في المواقع الالكترونية، حيث تعتبر الحاجة إلى وجود طريقة لتحديد مصداقية المعلومات عبر الانترنت أمراً ضرورياً خاصة وأنه يمكن لبعض من هذه المعلومات أن تكون غير موثوقة في مضمونها. سيقوم الباحث بتحليل كامل لتلك العوامل المؤثرة على ثقة المستخدمين بالمعلومات التي تحتويها المواقع الالكترونية، ومن هذا المنطلق قام الباحث بإعداد هذه الاستبانة كمتطلب لرسالة الماجستير في المعلوماتية لكلية الدراسات العليا والبحث العلمي في جامعة بوليتكنك فلسطين.

سيعرض الإستبيان مجموعة من المواقع التي تحتوي كل منها على معلومة ما بداخلها حيث يطلب من كل شخص قراءتها وابداء نسبة ثقته بما قرائه في هذا الموقع.

يشير الباحث إلى أن المعلومات الواردة في هذه الاستبانة لن تستخدم إلا لأغراض البحث العلمي فقط، لذا نرجو منكم التعاون معنا في تعبئة هذه الاستبانة بصدق وأمانة ونشكر لكم حسن تعاونكم معنا.

الباحث : م. رامي علي يوسف محسن.
مشرف الرسالة: د. عـادل الطـويل.

Required *

القســـــــم الأول: معلومات عامة

الجنس: *

- ذكر
- أنثى

كم سنة وأنت في مجال عملك الحالي؟ *

- أقل من سنتين
- من 2 الى 5
- من 6 الى 10
- أكثر من 10 سنوات

ما هو مجال تعليمك الأكاديمي؟؟ *

- Computer engineering/ IT
- Management/Financial/Accounting
- Engineering

13/.

استبيان حول الثقة بمحتوى المعلومات في المواقع الالكترونية

Other: ( )

القســـــــم الثاني : الثقة بمحتوى المعلومات في المواقع الالكترونية:

المـوقع الأول

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.ezinebase.net/automotive/motorcycles/seek-the-correct-label-for-your-company-2/ *

Seek The Correct Label For Your Company...

المـوقع الثاني

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.bitterskank.com/business_and_economy/business_resources/ *

Business Support (Paragraph on this site) "فقرة Business Support"

المـوقع الثالث

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://digiupdates.com/2013/bluetooth-handset-gloves-for-phone-calls/ *

Bluetooth Handset Gloves for Phone Calls

المـوقع الرابع

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.pbs.org/newshour/rundown/2013/03/how-good-is-your-hospital-depends-who-you-ask.html *

How Good Is Your Hospital? Depends Who You Ask

المـوقع الخامس

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://acnetreatmentzone.org/acne-medicines/acne-oral-medicines.html *

https://docs.google.com/forms/d/1PnieVfHI-eJ4qxecJQmZ3vLsP7HbaW9zxy1qvcVtl24/viewform

Acne Oral Medicines

المـوقع السادس

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.pmpnews.com/news/bd-purchases-cato-software-solutions *
BD Purchases Cato Software Solutions

المـوقع السابع

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.askq8.com/communication/communication-equipments/ *
Kuwait Communication Equipments

المـوقع الثامن

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.prweb.com/releases/World-Of-Brass/Door-Knobs/prweb10531544.htm *
New Door and Window Ranges from World of Brass

المـوقع التاسع

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.adverticia.com/three-types-of-parents-that-make-teachers-cringe/ *
Three Types of Parents That Make Teachers Cringe

المـوقع العاشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.thestreet.com/story/11871086/1/samsung-unloads-on-apple-tech-weekly.html *
Samsung Unloads on Apple: Tech Weekly

13/.

استبيان حول الثقة بمحتوى المعلومات في المواقع الالكترونية

المـوقع الحادي عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://soft.tecdiary.org/windows-applications/download-notepad-6-0-offline-installer/ *

Download Notepad++ 6.0 Offline Installer

المـوقع الثاني عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.cakebaker.co.uk/bake-traditional-florentines.html *

Bake Traditional Florentines

المـوقع الثالث عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://voterlux.com/story.php?title=tips-to-help-you-get-set-up-with-home-security *

Tips To Help You Get Set Up With Home Security

المـوقع الرابع عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://activegolftravel.com/395258-Seven-Secrets-for-Successful-Travel.html *

Seven Secrets for Successful Travel

المـوقع الخامس عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.chefosama.com/blog_ar.php?id=170#.UUe8G9ano-U *

فن الشواء

Submit

.Never submit passwords through Google Forms

https://docs.google.com/forms/d/1PnieVfHI-eJ4qxecJQmZ3vLsP7HbaW9zxy1qvcVtI24/viewform

Group #2 Questionnaire

13/ استبيان حول الثقة بمحتوى المعلومات في المواقع الالكترونية

Edit this form

استبيان حول الثقة بمحتوى المعلومات في المواقع الالكترونية

عزيزي القارئ/ عزيزتي القارئة:
يقوم الباحث بإجراء دراسة لمعرفة العوامل الأساسية المؤثرة في الثقة بمحتوى المعلومات الموجودة في المواقع الالكترونية، حيث تعتبر الحاجة إلى وجود طريقة لتحديد مصداقية المعلومات عبر الانترنت أمراً ضرورياً خاصة وأنه يمكن لبعض من هذه المعلومات أن تكون غير موثوقة في مضمونها. سيقوم الباحث بتحليل كامل لتلك العوامل المؤثرة على ثقة المستخدمين بالمعلومات التي تحتويها المواقع الالكترونية، ومن هذا المنطلق قام الباحث بإعداد هذه الاستبانة كمتطلب لرسالة الماجستير في المعلوماتية لكلية الدراسات العليا والبحث العلمي في جامعة بوليتكنك فلسطين.

سيعرض الإستبيان مجموعة من المواقع التي تحتوي كل منها على معلومة ما بداخلها حيث يطلب من كل شخص قراءتها وابداء نسبة ثقته بما قرائه في هذا الموقع.

يشير الباحث إلى أن المعلومات الواردة في هذه الاستبانة لن تستخدم إلا لأغراض البحث العلمي فقط، لذا نرجو منكم التعاون معنا في تعبئة هذه الاستبانة بصدق وأمانة ونشكر لكم حسن تعاونكم معنا.

الباحث : م. رامي علي يوسف محسن.
مشرف الرسالة: د. عـادل الطـويل.

Required *

القســـــــم الأول: معلومات عامة

الجنس: *
- ذكر
- أنثى

كم سنة وأنت في مجال عملك الحالي؟ *
- أقل من سنتين
- من 2 الى 5
- من 6 الى 10
- أكثر من 10 سنوات

ما هو مجال تعليمك الأكاديمي؟؟ *
- Computer engineering/ IT
- Management/Financial/Accounting
- Engineering

https://docs.google.com/forms/d/1bpVpl1rFsV76P-jUoPpbrL8ZqAbMcr-FQgsmk5t8aRs/viewform

13/ استبيان حول الثقة بمحتوى المعلومات في المواقع الالكترونية

Other: ____

القســـــــم الثاني : الثقة بمحتوى المعلومات في المواقع الالكترونية:

المـوقع الأول

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.gallup.com/poll/161399/10th-anniversary-iraq-war-mistake.aspx *
On 10th Anniversary, 53% in U.S. See Iraq War as Mistake

المـوقع الثاني

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.gadyd.com/2013/03/blog-post_1452.html *
عشرة أسباب لتقدم اليابان

المـوقع الثالث

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.ac-web.org/forums/showthread.php?177694-How-to-give-proper-CC&s=c7fecfa9dc454b021ffc90da1c90c793 *
how to give proper constructive criticism

المـوقع الرابع

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://linksheaven.co.uk/americanvisionwindows.com-520.html *
Window replacement phoenix

المـوقع الخامس

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.megeb.com/5-reasons-your-company-needs-to-hire-a-web-designer/ *
5 Reasons Your Company Needs To Hire A Web Designer

https://docs.google.com/forms/d/1bpVpl1rFsV76P-jUoPpbrL8ZqAbMcr-FQgsmk5t8aRs/viewform

المـوقع السادس

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://alwakad.net/dimofinf/news.php?action=show&id=17380 *

نجاح أول طائرة من دون طيار صنع الامارات العربية

المـوقع السابع

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://beirutspring.com/blog/2013/03/11/the-big-denial/ *

The Big Denial

المـوقع الثامن

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://reefbuilders.com/2013/01/19/rapture-deep-elusive-narcosis-angelfish/ *

Rapture of the Deep - The elusive Narcosis Angelfish

المـوقع التاسع

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.addatoday.com/2013/03/amitabh-bachchan-gives-autograph-to.html *

Amitabh Bachchan gives Autograph to Shahrukh Khan

المـوقع العاشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://thedemocraticdaily.com/2013/03/13/obama-losing-sequester-fight-dnc-pushes-back-hard/ *

Obama Losing Sequester Fight? DNC Pushes Back Hard

13/ استبيان حول الثقة بمحتوى المعلومات في المواقع الالكترونية

المـوقع الحادي عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.amazinginfographics.com/the-british-secret-for-a-good-nights-sleep/ *
The British Secret for a Good Night's Sleep

المـوقع الثاني عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.efytimes.com/e1/fullnews.asp?edid=102528 *
ASUS Working On Intel-Powered Smartphone: Report

المـوقع الثالث عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://masjidma.com/2013/03/05/member-of-anti-islamic-dutch-party-converts-to-islam/ *
Member of anti-Islamic Dutch party converts to Islam

المـوقع الرابع عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://www.dressesonlinesale.co.uk/dark-navy-stylish-sweetheart-a-line-dress.html *
Black A-line Floor-length Sweetheart Dress

المـوقع الخامس عشر

الرجاء تعبئة النسبة المئوية لثقتك في المربع (بدون %) بعد قرائتك لها

http://basementrejects.com/review/the-sixth-sense-1999/ *
The Sixth Sense (1999)

Submit

.Never submit passwords through Google Forms

https://docs.google.com/forms/d/1bpVpl1rFsV76P-jUoPpbrL8ZqAbMcr-FQgsmk5t8aRs/viewform

Appendix E

Appendix E: Abbreviations For Websites Articles URL Links (Evaluation)

Group 1 Selected Website Articles

website	Abbreviation
http://www.gallup.com/poll/161399/10th-anniversary-iraq-war-mistake.aspx	Ws1g1*
http://www.gadyd.com/2013/03/blog-post_1452.html	Ws2g1
http://www.ac-web.org/forums/showthread.php?177694-How-to-give-proper-CC&s=c7fecfa9dc454b021ffc90da1c90c793	Ws3g1
http://linksheaven.co.uk/americanvisionwindows.com-520.html	Ws4g1
http://www.megeb.com/5-reasons-your-company-needs-to-hire-a-web-designer/	Ws5g1
http://alwakad.net/dimofinf/news.php?action=show&id=17380	Ws6g1
http://beirutspring.com/blog/2013/03/11/the-big-denial/	Ws7g1
http://reefbuilders.com/2013/01/19/rapture-deep-elusive-narcosis-angelfish/	Ws8g1
http://www.addatoday.com/2013/03/amitabh-bachchan-gives-autograph-to.html	Ws9g1
http://thedemocraticdaily.com/2013/03/13/obama-losing-sequester-fight-dnc-pushes-back-hard/	Ws10g1
http://www.amazinginfographics.com/the-british-secret-for-a-good-nights-sleep/	Ws11g1
http://www.efytimes.com/e1/fullnews.asp?edid=102528	Ws12g1
http://masjidma.com/2013/03/05/member-of-anti-islamic-dutch-party-converts-to-islam/	Ws13g1
http://www.dressesonlinesale.co.uk/dark-navy-stylish-sweetheart-a-line-dress.html	Ws14g1
http://basementrejects.com/review/the-sixth-sense-1999/	Ws15g1

** Ws1g1: means the first website in group 1*

Group 2 Selected Website Articles

Website	Abbreviation
http://www.ezinebase.net/automotive/motorcycles/seek-the-correct-label-for-your-company-2/	Ws1g2*
http://www.bitterskank.com/business_and_economy/business_resources/	Ws2g2
http://digiupdates.com/2013/bluetooth-handset-gloves-for-phone-calls/	Ws3g2
http://www.pbs.org/newshour/rundown/2013/03/how-good-is-your-hospital-depends-who-you-ask.html	Ws4g2
http://acnetreatmentzone.org/acne-medicines/acne-oral-medicines.html	Ws5g2
http://www.pmpnews.com/news/bd-purchases-cato-software-solutions	Ws6g2
http://www.askq8.com/communication/communication-equipments/	Ws7g2
http://www.prweb.com/releases/World-Of-Brass/Door-Knobs/prweb10531544.htm	Ws8g2
http://www.adverticia.com/three-types-of-parents-that-make-teachers-cringe/	Ws9g2
http://www.thestreet.com/story/11871086/1/samsung-unloads-on-apple-tech-weekly.html	Ws10g2
http://soft.tecdiary.org/windows-applications/download-notepad-6-0-offline-installer/	Ws11g2
http://www.cakebaker.co.uk/bake-traditional-florentines.html	Ws12g2
http://voterlux.com/story.php?title=tips-to-help-you-get-set-up-with-home-security	Ws13g2
http://activegolftravel.com/395258-Seven-Secrets-for-Successful-Travel.html	Ws14g2
http://www.chefosama.com/blog_ar.php?id=170#.UUe8G9ano-U	Ws15g2

** Ws1g2: means the first website in group 2*

Selected Medical Website Articles

Website	Abbreviation
http://health.yahoo.net/health/breast-cancer-symptoms	ws1
http://breast-cancer-symptoms.blogs2day.net/some-symptoms-of-breast-cancer/	ws2
http://www.webmd.com/breast-cancer/guide/understanding-breast-cancer-symptoms	ws3
http://www.bresca.org/symptoms-of-breast-cancer-in-women.html	ws4
http://eb.gmnews.com/news/2012-10-02/Front_Page/Symptoms_of_breast_cancer.html	ws5
http://www.ibcsupport.org/pictures/vsc.html	ws6
http://www.davidgreenspunmd.com/breast_cancer/symptoms/	ws7
http://www.nlm.nih.gov/medlineplus/breastcancer.html	ws8
http://mazumdarshawcancercenter.wordpress.com/2011/09/23/breast-cancer-symptoms-prevention-treatmen/	ws9
http://formabest.blogspot.com/2012/09/breast-cancer-symptoms-and-signs.html	ws10
http://www.lifeandtrend.com/health/cancer/What-to-Know-About-Breast-Cancer/	ws11
http://www.caring.com/articles/breast-cancer-risk	ws12
http://www.onhealth.com/breast_cancer/page4.htm#what_are_breast_cancer_symptoms_and_signs	ws13
http://www.drugs.com/health-guide/breast-cancer.html	ws14
http://blog-forbreastcancer.blogspot.com/2010/04/breast-cancer-symptoms.html	ws15
http://www.everydayhealth.com/breast-cancer/understanding/warning-signs.aspx	ws16
http://www.resc.ca/breast-cancer-and-its-symptoms.html	ws17
http://www.healthplus24.com/womens-health/breast-cancer/symptoms-of-breast-cancer.aspx	ws18
http://symptoms-ofbreastcancer.blogspot.com	ws19
http://www.wellsphere.com/breast-cancer-article/signs-symptoms-of-breast-cancer/116213	ws20

Appendix F

Appendix F: Hotelling's T^2 Test Before and After Excluding High Error Webpages (Evaluation)

- *Hotelling's T^2:*

We used hotelling's T^2 paradigm to test whether the CTA agent answers vector (P x 1) μ_0 is a plausible value for the respondents' population answers mean vector μ. Hence, the squared distance is:

$$T^2 = (\bar{X} - \mu_0)' \left(\frac{S}{n}\right)^{-1} (\bar{X} - \mu_0) = n(\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0)$$

This is called the Hotelling's T^2, where: $\bar{X} = \frac{1}{n}\sum_i X_i$, $S = \frac{1}{n-1}\sum_i (X_i - \bar{X})\,(X_i - \bar{X})'$

If the observed T^2 value is "large" we reject H_0: $\mu = \mu_0$, if the question is how to decide how large is large, we need the sampling distribution of T^2 when the hypothesized mean vector is correct:

$$T^2 = \frac{(n-1)p}{(n-p)} F_{p,n-p}$$

We reject the null hypothesis H_0: $\mu = \mu_0$ for the p-dimensional vector μ at level α when:

$$T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

Where $F_{p,n-p}(\alpha)$ is the upper percentile of the central *F* distribution with *p* and *n-p* degrees of freedom.

- *Hotelling's T^2 Analysis:*

As we mentioned previously we will reject the null hypothesis H_0: $\mu = \mu_0$ for the p-dimensional vector μ at level α when:

$$T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

Which is equivalent to:

$$n(\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

In our evaluation experiment, we have for each group *n=20* respondents and *p=15* webpages while we select level α=0.05 then $F_{p,n-p}(\alpha) = F_{15,5}(0.05) = 4.62$ as shown in figure F.1 that

represents upper percentage points of the F distribution and show how we extract the value of $F_{15,5}(0.05)$.Hence, the equation after substitution *for each group* will be:

$$20(\bar{X} - \mu_0)' S^{-1}(\bar{X} - \mu_0) > \frac{(20-1)15}{(20-15)} \times 4.62$$

$$\rightarrow 20(\bar{X} - \mu_0)' S^{-1}(\bar{X} - \mu_0) > 263.34$$

Now, if the left side of equation is greater than 263.34, we reject the hypothesis H_0: $\mu = \mu_0$ for the p-dimensional respondents mean vector μ at level α=0.05. We can derive $(\bar{X} - \mu_0)'$ from table 7.3 for group1 and 7.4 for group2 as shown in table F.1 and table F.2 respectively. $(\bar{X} - \mu_0)'$ Is a vector with a dimension of *1 × p*

TABLE F.1 : $(\bar{X} - \mu_0)'$ VECTOR VALUES FROM RESPONDENTS MEAN VECTOR AND AGENT ANSWERS VECTOR/GROUP1

For Group 1	Ws1g1	Ws2g1	Ws3g1	Ws4g1	Ws5g1	Ws6g1	Ws7g1	Ws8g1
Respondents Mean ($\bar{X}$)	81.9	61.25	44.75	46	67.65	51.65	61.35	69.1
Agent Answers (μ0)	84.872	55.0817	39.2911	38.9955	61.3115	60.0012	70.5667	71.1055
($\bar{X}$ - μ0)'	**-2.972**	**6.1683**	**5.4589**	**7.0045**	**6.3385**	**-8.3512**	**-9.2167**	**-2.0055**
For Group 1	**Ws9g1**	**Ws10g1**	**Ws11g1**	**Ws12g1**	**Ws13g1**	**Ws14g1**	**Ws15g1**	
Respondents Mean ($\bar{X}$)	55.4	66.85	69.4	68.4	73.55	67.4	60.15	
Agent Answers (μ0)	52.0811	78.662	71.0317	81.742	44.3887	71.742	50.9521	
($\bar{X}$ - μ0)'	**3.3189**	**-11.812**	**-1.6317**	**-13.342**	**29.1613**	**-4.342**	**9.1979**	

TABLE F.2:$(\bar{X} - \mu_0)'$ VECTOR VALUES FROM RESPONDENTS MEAN VECTOR AND AGENT ANSWERS VECTOR/GROUP2

For Group 2	Ws1g2	Ws2g2	Ws3g2	Ws4g2	Ws5g2	Ws6g2	Ws7g2	Ws8g2
Respondents Mean ($\bar{X}$)	76.85	46.5	60.55	79	53.5	69.5	42.4	75.8
Agent Answers (μ0)	77.973	51.0921	60.9847	87.062	61.0787	78.142	33.0993	77.0823
($\bar{X}$ - μ0)'	**-1.123**	**-4.5921**	**-0.4347**	**-8.062**	**-7.5787**	**-8.642**	**9.3007**	**-1.2823**
For Group 2	**Ws9g2**	**Ws10g2**	**Ws11g2**	**Ws12g2**	**Ws13g2**	**Ws14g2**	**Ws15g2**	
Respondents Mean ($\bar{X}$)	68.35	71.5	45.8	61.75	44.5	68.3	72.75	
Agent Answers (μ0)	60.98071	77.692	32.8515	77.0973	34.1417	60.9736	61.0682	
($\bar{X}$ - μ0)'	**7.36929**	**-6.192**	**12.9485**	**-15.3473**	**10.3583**	**7.3264**	**11.6818**	

df for numerator N_1													df for denominator N_2
15	20	24	30	40	50	60	100	120	200	500	∞	Pr	
9.49	9.58	9.63	9.67	9.71	9.74	9.76	9.78	9.80	9.82	9.84	9.85	.25	
61.2	61.7	62.0	62.3	62.5	62.7	62.8	63.0	63.1	63.2	63.3	63.3	.10	1
246	248	249	250	251	252	252	253	253	254	254	254	.05	
3.41	3.43	3.43	3.44	3.45	3.45	3.46	3.47	3.47	3.48	3.48	3.48	.25	
9.42	9.44	9.45	9.46	9.47	9.47	9.47	9.48	9.48	9.49	9.49	9.49	.10	2
19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	.05	
99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	.01	
2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47	2.47	2.47	2.47	2.47	.25	
5.20	5.18	5.18	5.17	5.16	5.15	5.15	5.14	5.14	5.14	5.14	5.13	.10	3
8.70	8.66	8.64	8.62	8.59	8.58	8.57	8.55	8.55	8.54	8.53	8.53	.05	
26.9	26.7	26.6	26.5	26.4	26.4	26.3	26.2	26.2	26.2	26.1	26.1	.01	
2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	.25	
3.87	3.84	3.83	3.82	3.80	3.80	3.79	3.78	3.78	3.77	3.76	3.76	.10	4
5.86	5.80	5.77	5.75	5.72	5.70	5.69	5.66	5.66	5.65	5.64	5.63	.05	
14.2	14.0	13.9	13.8	13.7	13.7	13.7	13.6	13.6	13.5	13.5	13.5	.01	
1.89	1.88	1.88	1.88	1.88	1.88	1.87	1.87	1.87	1.87	1.87	1.87	.25	
3.24	3.21	3.19	3.17	3.16	3.15	3.14	3.13	3.12	3.12	3.11	3.10	.10	
4.62	4.56	4.53	4.50	4.46	4.44	4.43	4.41	4.40	4.39	4.37	4.36	.05	5
9.72	9.55	9.47	9.38	9.29	9.24	9.20	9.13	9.11	9.08	9.04	9.02	.01	
1.76	1.76	1.75	1.75	1.75	1.75	1.74	1.74	1.74	1.74	1.74	1.74	.25	
2.87	2.84	2.82	2.80	2.78	2.77	2.76	2.75	2.74	2.73	2.73	2.72	.10	
3.94	3.87	3.84	3.81	3.77	3.75	3.74	3.71	3.70	3.69	3.68	3.67	.05	6
7.56	7.40	7.31	7.23	7.14	7.09	7.06	6.99	6.97	6.93	6.90	6.88	.01	
1.68	1.67	1.67	1.66	1.66	1.66	1.65	1.65	1.65	1.65	1.65	1.65	.25	
2.63	2.59	2.58	2.56	2.54	2.52	2.51	2.50	2.49	2.48	2.48	2.47	.10	
3.51	3.44	3.41	3.38	3.34	3.32	3.30	3.27	3.27	3.25	3.24	3.23	.05	7
6.31	6.16	6.07	5.99	5.91	5.86	5.82	5.75	5.74	5.70	5.67	5.65	.01	
1.62	1.61	1.60	1.60	1.59	1.59	1.59	1.58	1.58	1.58	1.58	1.58	.25	
2.46	2.42	2.40	2.38	2.36	2.35	2.34	2.32	2.32	2.31	2.30	2.29	.10	
3.22	3.15	3.12	3.08	3.04	2.02	3.01	2.97	2.97	2.95	2.94	2.93	.05	8
5.52	5.36	5.28	5.20	5.12	5.07	5.03	4.96	4.95	4.91	4.88	4.86	.01	
1.57	1.56	1.56	1.55	1.55	1.54	1.54	1.53	1.53	1.53	1.53	1.53	.25	
2.34	2.30	2.28	2.25	2.23	2.22	2.21	2.19	2.18	2.17	2.17	2.16	.10	
3.01	2.94	2.90	2.86	2.83	2.80	2.79	2.76	2.75	2.73	2.72	2.71	.05	9
4.96	4.81	4.73	4.65	4.57	4.52	4.48	4.42	4.40	4.36	4.33	4.31	.01	

Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 18, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.

FIGURE F.1 : UPPER PERCENTAGE POINTS OF THE F DISTRIBUTION

S^{-1} in the equation represents the inverse of S "sample covariance matrix n samples" that has $p \times p$ dimension, where $S = \frac{1}{n-1}\sum_i (X_i - \bar{X})\,(X_i - \bar{X})'$ the table F.3 and table F.4 show the results for calculating S^{-1} for both groups.

By calculating S^{-1} for each group, we can substitute all variable in our hypothesis equation to check if H_0 reject or not, so:

Group#1:

$$\rightarrow 20(\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0) > 263.34 \quad \rightarrow \mathbf{1189.13} > 263.34 \text{ (Rejected)}$$

Group#2:

$$\rightarrow 20(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0) > 263.34 \quad \rightarrow \mathbf{193.02} >^{?} 263.34 \text{ (Accepted)}$$

We will ***reject*** the hypothesis H_0: μ = μ_0 for the p-dimensional respondents mean vector μ at level α=0.05 for group1 as we ***accept*** the hypothesis H_0 for group2. In the next section we will study the hotelling's T^2 after *excluding* the webpages missed by their confidence intervals in group1.

- *Hotelling's T^2 Analysis Excluding Missed Webpage For Group1:*

In this section we will exclude the missed web-pages that exceeded their confidence intervals in group1 and see how the result if hotelling's T^2 test will give us. As mentioned previously, three web-pages are missed by our trust model in group1 (*Ws10g1, Ws12g1 and Ws13g1*) that will make p=12 and $F_{p,n-p}(\alpha) = F_{12,8}(0.05) = 3.28$, after we exclude them the hotelling's T^2 analysis will be:

$$20(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0) > \frac{(20-1)12}{(20-12)} \times 3.28$$

$$\rightarrow 20(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0) > 93.48$$

And $(\bar{X}-\mu_0)'$, S^{-1} values will be as shown in table F.5 and F.6 respectively:

$$\rightarrow \mathbf{32.26} >^{?} 93.48 \text{ NO (Accepted)}$$

So, we ***accept*** the hypothesis H_0: μ = μ_0 for the p-dimensional respondents mean vector μ at level α=0.05 for group#1 after the exclusion of the missed web-pages by the confidence interval. The existence of these web-pages in hotelling's T^2 calculation makes the hypothesis H_0 rejected since their difference between respondents mean and agent answer is large as it appears in table 7.3.

As mentioned before, these large differences are due to the limitations of our model to handle some factors on its calculation that effect on its judgment precision.

TABLE F.3: S^{-1} FOR GROUP1 DATASET

	Ws1g1	Ws2g1	Ws3g1	Ws4g1	Ws5g1	Ws6g1	Ws7g1	Ws8g1	Ws9g1	Ws10g1	Ws11g1	Ws12g1	Ws13g1	Ws14g1	Ws15g1
Ws1g1	0.089482	-0.01803	0.027324	0.030133	0.015296	-0.00057	-0.02095	-0.01493	0.002057	-0.04027	0.0177	-0.01443	0.022191	-0.00223	0.008401
Ws2g1	-0.01803	0.012905	-0.01569	0.000361	-0.01376	0.008337	0.005939	-0.0021	-0.00134	0.017585	-0.00275	0.004591	-0.00647	-0.00073	0.002116
Ws3g1	0.027324	-0.01569	0.035003	-0.00431	0.020496	-0.01744	-0.01242	0.008222	0.002656	-0.03196	0.008284	-0.00982	0.01106	0.007892	-0.00625
Ws4g1	0.030133	0.000361	-0.00431	0.023018	-0.00419	0.010353	-0.0036	-0.01403	-0.00141	0.000297	0.009321	-0.00427	0.001593	-0.00431	0.007632
Ws5g1	0.015296	-0.01376	0.020496	-0.00419	0.02432	-0.01628	-0.0078	0.005526	0.000827	-0.02685	0.005116	-0.00639	0.005022	0.003743	-0.00371
Ws6g1	-0.00057	0.008337	-0.01744	0.010353	-0.01628	0.020222	0.003958	-0.00926	-0.00112	0.020315	-0.00343	0.002825	-0.00456	-0.00684	0.00548
Ws7g1	-0.02095	0.005939	-0.01242	-0.0036	-0.0078	0.003958	0.011899	-0.00153	-0.0004	0.016194	-0.00439	0.003148	-0.00767	-0.00523	-0.00065
Ws8g1	-0.01493	-0.0021	0.008222	-0.01403	0.005526	-0.00926	-0.00153	0.017405	-0.00162	-0.00632	-0.0063	0.000836	0.000779	0.008328	-0.00494
Ws9g1	0.002057	-0.00134	0.002656	-0.00141	0.000827	-0.00112	-0.0004	-0.00162	0.004804	-0.00297	0.001405	8.32E-05	0.002281	-0.00361	-0.00166
Ws10g1	-0.04027	0.017585	-0.03196	0.000297	-0.02685	0.020315	0.016194	-0.00632	-0.00297	0.046739	-0.00927	0.009628	-0.0168	-0.00648	0.0019
Ws11g1	0.0177	-0.00275	0.008284	0.009321	0.005116	-0.00343	-0.00439	-0.0063	0.001405	-0.00927	0.016103	-0.00732	-0.00081	0.000945	0.000503
Ws12g1	-0.01443	0.004591	-0.00982	-0.00427	-0.00639	0.002825	0.003148	0.000836	8.32E-05	0.009628	-0.00732	0.011466	-0.00108	-0.00186	0.000527
Ws13g1	0.022191	-0.00647	0.01106	0.001593	0.005022	-0.00456	-0.00767	0.000779	0.002281	-0.0168	-0.00081	-0.00108	0.015182	0.001122	0.001334
Ws14g1	-0.00223	-0.00073	0.007892	-0.00431	0.003743	-0.00684	-0.00523	0.008328	-0.00361	-0.00648	0.000945	-0.00186	0.001122	0.015371	-0.00332
Ws15g1	0.008401	0.002116	-0.00625	0.007632	-0.00371	0.00548	-0.00065	-0.00494	-0.00166	0.0019	0.000503	0.000527	0.001334	-0.00332	0.007202

TABLE F.4: S^{-1} FOR GROUP2 DATASET

	Ws1g2	Ws2g2	Ws3g2	Ws4g2	Ws5g2	Ws6g2	Ws7g2	Ws8g2	Ws9g2	Ws10g2	Ws11g2	Ws12g2	Ws13g2	Ws14g2	Ws15g2
Ws1g2	0.070555	-0.01987	-0.00886	0.061154	-0.0283	0.024785	0.026133	-0.00082	-0.00692	-0.02066	0.024769	0.009851	-0.00227	0.031953	-0.05684
Ws2g2	-0.01987	0.021379	-0.00867	-0.02996	0.016652	-0.01087	-0.01087	0.003327	-0.00028	0.009252	-0.01344	-0.00043	-0.00054	-0.01169	0.019949
Ws3g2	-0.00886	-0.00867	0.018672	0.002056	-0.00312	-0.00498	-0.00225	-0.00398	0.007136	0.001267	0.002726	-0.00459	-0.00145	-0.00479	0.006589
Ws4g2	0.061154	-0.02996	0.002056	0.105792	-0.04286	0.020019	0.02342	0.000263	-0.01157	-0.02108	0.049069	0.007622	-0.00688	0.033484	-0.06838
Ws5g2	-0.0283	0.016652	-0.00312	-0.04286	0.040668	-0.02501	0.004112	0.002021	0.016163	0.012138	-0.05448	-0.00456	0.005723	-0.01269	0.033773
Ws6g2	0.024785	-0.01087	-0.00498	0.020019	-0.02501	0.029702	-0.00324	-0.0035	-0.01486	-0.01081	0.035671	0.000732	-0.00115	0.010722	-0.02253
Ws7g2	0.026133	-0.01087	-0.00225	0.02342	0.004112	-0.00324	0.03755	0.003229	0.016053	-0.00572	-0.03326	0.004834	0.004222	0.01533	-0.02295
Ws8g2	-0.00082	0.003327	-0.00398	0.000263	0.002021	-0.0035	0.003229	0.005586	-0.00063	0.001309	-0.00534	0.00172	-0.00028	0.000754	-0.00153
Ws9g2	-0.00692	-0.00028	0.007136	-0.01157	0.016163	-0.01486	0.016053	-0.00063	0.021824	0.003163	-0.03716	1.45E-05	0.003778	-0.00356	0.009644
Ws10g2	-0.02066	0.009252	0.001267	-0.02108	0.012138	-0.01081	-0.00572	0.001309	0.003163	0.010704	-0.01226	-0.0035	0.000167	-0.01112	0.018113
Ws11g2	0.024769	-0.01344	0.002726	0.049069	-0.05448	0.035671	-0.03326	-0.00534	-0.03716	-0.01226	0.103464	-9E-05	-0.01264	0.008011	-0.0325
Ws12g2	0.009851	-0.00043	-0.00459	0.007622	-0.00456	0.000732	0.004834	0.00172	1.45E-05	-0.0035	-9E-05	0.00855	0.000383	0.004223	-0.0083
Ws13g2	-0.00227	-0.00054	-0.00145	-0.00688	0.005723	-0.00115	0.004222	-0.00028	0.003778	0.000167	-0.01264	0.000383	0.004791	-0.00051	0.003518
Ws14g2	0.031953	-0.01169	-0.00479	0.033484	-0.01269	0.010722	0.01533	0.000754	-0.00356	-0.01112	0.008011	0.004223	-0.00051	0.020796	-0.02924
Ws15g2	-0.05684	0.019949	0.006589	-0.06838	0.033773	-0.02253	-0.02295	-0.00153	0.009644	0.018113	-0.0325	-0.0083	0.003518	-0.02924	0.059772

TABLE F.5:$(\bar{X} - \mu_0)'$VECTOR VALUES AFTER EXCLUDING *WS10G1, WS12G1 AND WS13G1*/GROUP1

For Group 1	Ws1g1	Ws2g1	Ws3g1	Ws4g1	Ws5g1	Ws6g1	Ws7g1	Ws9g1	Ws11g1	Ws14g1	Ws15g1
Respondents Mean ($\bar{X}$)	81.9	61.25	44.75	46	67.65	51.65	61.35	55.4	69.4	67.4	60.15
Agent Answers (μ0)	84.872	55.0817	39.2911	38.9955	61.3115	60.0012	70.5667	52.0811	71.0317	71.742	50.9521
($\bar{X}$ - μ0)'	**-2.972**	**6.1683**	**5.4589**	**7.0045**	**6.3385**	**-8.3512**	**-9.2167**	**3.3189**	**-1.6317**	**-4.342**	**9.1979**

TABLE F.6: S^{-1} FOR GROUP1DATASETAFTER EXCLUDING /GROUP1

	Ws1g1	Ws2g1	Ws3g1	Ws4g1	Ws5g1	Ws6g1	Ws7g1	Ws8g1	Ws9g1	Ws11g1	Ws14g1	Ws15g1
Ws1g1	0.040784	-0.00183	-0.00272	0.024562	-0.00361	0.012716	-0.00516	-0.01679	-0.00118	0.009227	-0.00698	0.007977
Ws2g1	-0.00183	0.006172	-0.00332	0.000819	-0.00377	0.000972	-0.00022	-3.2E-05	-0.00024	0.001168	0.001716	0.001478
Ws3g1	-0.00272	-0.00332	0.012028	-0.00571	0.002045	-0.00415	-0.00133	0.00472	0.000812	0.000225	0.003333	-0.00499
Ws4g1	0.024562	0.000819	-0.00571	0.020123	-0.00297	0.008597	-0.0032	-0.01235	-0.00141	0.007816	-0.00416	0.007035
Ws5g1	-0.00361	-0.00377	0.002045	-0.00297	0.006545	-0.00314	0.000551	0.001041	-0.00028	-0.00214	-0.00058	-0.00159
Ws6g1	0.012716	0.000972	-0.00415	0.008597	-0.00314	0.010088	-0.00246	-0.00547	-9.6E-05	0.000831	-0.00371	0.003959
Ws7g1	-0.00516	-0.00022	-0.00133	-0.0032	0.000551	-0.00246	0.005903	0.000275	0.000862	-0.00187	-0.00322	-0.00088
Ws8g1	-0.01679	-3.2E-05	0.00472	-0.01235	0.001041	-0.00547	0.000275	0.015587	-0.00193	-0.007	0.007314	-0.00424
Ws9g1	-0.00118	-0.00024	0.000812	-0.00141	-0.00028	-9.6E-05	0.000862	-0.00193	0.004438	0.00155	-0.00385	-0.00179
Ws11g1	0.009227	0.001168	0.000225	0.007816	-0.00214	0.000831	-0.00187	-0.007	0.00155	0.010273	-0.00099	0.001603
Ws14g1	-0.00698	0.001716	0.003333	-0.00416	-0.00058	-0.00371	-0.00322	0.007314	-0.00385	-0.00099	0.01431	-0.0028
Ws15g1	0.007977	0.001478	-0.00499	0.007035	-0.00159	0.003959	-0.00088	-0.00424	-0.00179	0.001603	-0.0028	0.006663