TABLE OF CONTENTS

Page No.

Statement of Originality………………………………………….………….. I

Acknowledgement……………………………………………………….…….II

Abstract………………………………………………………………………..III

Table of Contents ……………………………………………………………...IV

List of Figures…………………………………………………………………VIII

List of Tables…………………………………………………………………..X

Chapter 1 Introduction and Outline………………………………………....1

* 1. Introduction…………………………………………………………..1
  2. Research Objectives………………………………………………….5
  3. Proposed Framework………………………………………..............6
  4. Organization of Thesis……………………………………………….8
  5. Chapter Summary………………………………………….…………8

Chapter 2 Literature Review…………………………………………………9

2.1 Evolution of Web……………………………….……………………9

2.2 Web 2.0 services……………………………………………………..15

2.3 Collaboration vs. Participation……………………………………....26

2.3.1 Collaboration………………………………………………26

2.3.2 Participation………………………………………………..31

2.4 Virtual Community…………………………………………………..31

2.5 Recommender System…………………………………………….....35

2.5.1 Common Technologies…………………………………….36

2.5.2 Recommender System Process…………………………….39

2.5.2.1 Inputs for Recommender System……...............40

2.5.2.2 Recommendation Generation Process...............43

2.5.2.3 Outputs for Recommender System…................45

2.5.3 Types of Recommender System…………………………….45

2.5.4 Evaluation of Recommender System……………………….48

2.5.5 Classification of Recommender System…………………….51

2.6 Collaborative Filtering………………………………………………..55

2.6.1 Similarity Metrics……………………………………………61

2.6.1.1 Pearson Correlation Coefficient………………….61

2.6.1.2 Cosine Measure……………………………………62

2.6.1.3 Distance measure(s)……………………………...63

2.6.1.4 Jaccard (and Tanimoto) Coefficient……………..65

2.6.1.5 Tversky Coefficient………………………………66

2.6.1.6 Fractional Match Coefficient…………………….67

2.6.1.7 Compressed Bit Similarity (cbit)………………..68

2.6.2 Drawbacks of Collaborative Filtering……………………...71

2.7 Chapter Summary……………………………………………………..72

**Chapter 3 Proposed Framework……………………………………………...74**

3.1 The Proposed System…………………………………………………74

3.2 The System Architectural View………………………………………76

3.3 The COMREC System……………………………………………….78 3.3.1 Interest Mining Module…………………………………….79

3.3.1.1 Interest Vector…………………………………….81 3.3.1.2 Interest Similarity Score………………………….81

3.3.1.3 Collaborative Interest Group Construction……….82

3.3.2 Accessing Expertise in Collaborative Interest Group……...84

3.3.3 Recommendation Module…………………………………..85

3.3.3.1 Review Data Server……………………………….86 3.3.3.2 Sentiment Mining Engine……………………......87

3.3.3.3 User Interface Result Recommendation………….95

3.4 Chapter Summary……………………………………………………..95

Chapter 4 Experimental Results and Analysis………………………………96

4.1 Illustration…………………………………………………………….96 4.1.1 Interest Mining Module……………………………………...97

4.1.1.1 Interest Vector calculations ………………………..97

4.1.1.2 Interest Similarity Score calculations……………..100

4.1.1.3 Collaborative Interest Group construction…………101

4.1.2 Accessing Expertise in Collaborative Interest Group………..105

4.1.3 Recommendation Module…………………………………….107

4.1.3.1 Feature Extraction………………………………….107

4.1.3.2 Opinion Direction Identification…………………...112

4.1.3.3 Review Orientation…………………………………116

Chapter 5 Conclusion and Future Research………………………………...117

5.1 Research Summary……………………………………………….…..117

5.2 Limitations of the Study……………………………………………...119

5.3 Future Directions……………………………………………………..119

5.3.1 Taking the Temporal Opinion Quality (TOQ) into account….120

5.3.2 Optimizing the recommendation module algorithm…………120

5.3.2.1 Taking the implicit features of the product into

account………………………………………………120

5.3.2.2 Taking the strength of nouns into account………….120

5.3.3 Alternatives for calculating the Interest Similarity between

researchers…………………………………………………….120

5.3.4 An alternative to Expert Calculation………………………….121

5.3.5 An alternative to find Adjective Polarity……………………..121

5.4 Conclusion……………………………………………………...........121

References……………………………………………………………………….123

Appendix A: POS Tagging……………………………………………………..127

Appendix B: Stemming………………………………………………………...129

Appendix C: WordNet…………………………………………………………146

Appendix D: Publication……………………………………………………….149

**LIST OF FIGURES**

Figure 1.1: Timeline depicting the evolution of Web………………………………………...1

Figure 1.2:Web 2.0 Technologies……………………………………………………………3

Figure 2.1:Web-based 'collaboration-ware'………………………………………………….11

Figure 2.2:A Simple Picture Web Evolution………………………………………………...11

Figure 2.3: Web 2.0 Framework……………………………………………………………...17

Figure 2.4: Web 2.0 Definitions……………………………………………………………...17

Figure 2.5: Web 2.0 Landscape………………………………………………………………18

Figure 2.6:‘Facebook’ and ‘Orkut’- The best-known social networking sites………………19

Figure 2.7:Blogging - The most recognized example of web 2.0……………………………20

Figure 2.8:Wikipedia- A collaborative dictionary being edited in real time by anyone…….21

Figure 2.9:DEL.ICIO.US - An example of a site that uses a folksonomy to organize

bookmarks………………………………………………………………………...23

Figure 2.10:Flickr- A website which combines a social network with user generated

content and used to collaborate on photo projects ……………………………...25

Figure 2.11:Skype **-** The best-known service provider………………………………………26

Figure 2.12:Different approaches to dealing with an issue………………………………….29

Figure 2.13: Generic Input—Process—Output Model of Recommender Systems ……...….40

Figure 2.14: Product Taxonomy for personal care and grooming category……………….....42

Figure 2.15: Sample Association Rules Mined from Table 2.5………………………………44

Figure 2.16: Examples of Book and Movie Recommender Systems…………………………46

Figure 2.17: Possible Recommendations……………………………………………………..49

Figure 2.18: Product Matrix……………………………………………………………….…53

Figure 2.19: Binary Values for Users A and B……………………………………………....65

Figure 2.20: Sample Product Taxonomy for Consumer Products…………………………...70

Figure 3.1: The System Architecture………………………………………………………..77

Figure 3.2: Collaborative Interest Group…………………………………………………...79

Figure 3.3: Recommender Systems as a solution to Information overload………………...85

Figure 3.4**:** The WordNet (<http://wordnet.princeton.edu/>)...................................................92

**LIST OF TABLES**

Table 2.1: Differences between Web 1.0, Web 2.0 and Web 3.0……………………………..15

Table 2.2:Differences between Web 2.0 communities (w2.0c) and Virtual Community……35

Table 2.3: Types of Input Data Used in Recommender Systems……………………………..41

Table 2.4: Product Database for Color TVs…………………………………………………..43

Table 2.5: Customer Product Purchase Database……………………………………………..44

Table 2.6: Customer Ratings of Web Pages on a Scale of 1-7………………………………..58

Table 2.7: Similarity Values Computed for the Target User C3 ……………………………...59

Table 2.8: Prediction Scores for Web Pages I2 and I7 ………………………………………...60

Table 4.1: Sample blog entries of 5 researchers................................................................97