Informatics Institute of Technology

In collaboration with

University of Westminster, UK.

“**Software Agent Based News Recommender System for an Online News Environment”**

A Dissertation By

Sriyal Himesh Jayasinghe

Supervised by

Ms. Naomi Krishnaraja

Submitted in partial fulfilment of the requirements for the

M.Sc in Advanced Software Engineering

Department of Computing

**August 2014**

©The copyright for this project and all its associated products resides with Informatics Institute of Technology

|  |  |
| --- | --- |
| Declaration I hereby certify that this project report and all the artefacts associated with it is my **own** work and it has not been submitted before nor is currently being submitted for any degree programme.  Full Name of Student: Jayasinghe Arachchilage Sriyal Himesh Jayasinghe  Registration No : 2009077 | |
| Signature : ................................................... | Date : ........................................ |

# Abstract

News has grown to be an essential in the contemporary world and online news media plays a major role in the news domain by publishing the latest news and allowing the readers to access them from any place at any time. Online news companies generally sort news articles according to the time it is published and due to the high volume of news articles and instance updates most online readers face a limitation in finding news articles which would match their specific preferences. The only facility provided by most online news domains is to overcome this limitation is to provide a searching facility for the readers to manually search for preferred news articles which would not impress all readers due to the time and the effort that has to be invested in the searching process. Previous work done in the field of news recommendation has only looked into providing a solution either supporting a web based presentation or a mobile based presentation without looking in to the prospect of providing one presentation layer for both the mediums. Another drawback of the current solutions is not considering the scalability of the system as a key success factor which would have a negative impact on the performance of the system due to the vast amount of articles and readers of online news environments.

To overcome the above mentioned limitations of online news environments, NewsRec system was introduced which uses a hybrid recommender engine to achieve a high accuracy in the recommendation. NewsRec system was designed and implemented to have only one presentation layer to support any type of a device and platform and software agents were utilized to induce better collation and communication between different activities of the recommender system thus making the scalability of the system to increase.

All agents were implemented using JADE agent framework and a RESTful API exposed through NodeJS and Redis server were used to read and store data coming through the clients and Java Persistence and Hibernate to store the processed data. Implemented system was tested thoroughly under different conditions and the NewsRec system was evaluated by evaluators of various domains. Eventually, the test results attested that the analysis, design, implementation and documentation have been carried out in an effective and in an efficient manner.

Subject Descriptors:

H.3.3 Information Search and Retrieval

I.2.11 Multiagent systems

Key Words:

Information Filtering, Personalization, JADE framework

# Acknowledgement

The effort that was put by me in to completing this project has been a very tough expedition throughout but at the same time it was a wonderful experience. However, it would not have been possible without the kind support, assistance and courage of many people who stood behind me. I would like to extend my sincere thanks to all of them.

I am highly indebted to my supervisor Ms. Naomi Krishnarajah, Head of the Department of Computer Science, Informatics Institute of Technology for her constant guidance and supervision as well as for encouragement and ideas given me to strive to greater heights and to achieve my fullest potential.

I hereby express my gratitude to Prof. Alfred Perera, Dean of the Informatics Institute of Technology and the management of Informatics Institute of Technology for providing me the opportunity for engaging this work.

I would like to express my special gratitude and thanks to everyone who assisted me by participating online surveys, interviews and contributing their ideas for requirements gathering and critical evaluation.

Finally I would like to express my great gratitude towards my wife for her kind co-operation and encouragement which helped me to complete the project successfully.

# Table of Contents

[Declaration ii](#_Toc396716496)

[Abstract ii](#_Toc396716497)

[Acknowledgement iii](#_Toc396716498)

[Table of Contents iv](#_Toc396716499)

[List of Figures viii](#_Toc396716500)

[List of Tables viii](#_Toc396716501)

[List of Abbreviations xi](#_Toc396716502)

[Chapter 1: Introduction 1](#_Toc396716503)

[1.1 Chapter Overview 1](#_Toc396716504)

[1.2 Problem Domain 1](#_Toc396716505)

[1.3 Previous Work 2](#_Toc396716506)

[1.4 Project Aim 3](#_Toc396716507)

[1.5 Project Objectives 4](#_Toc396716508)

[1.6 Feature of the Prototype 6](#_Toc396716509)

[1.7Resource Requirements 6](#_Toc396716510)

[1.8 Project Document Structure 7](#_Toc396716511)

[Chapter 2: Literature Review 8](#_Toc396716512)

[2.1 Chapter Overview 8](#_Toc396716513)

[2.2 Recommendation Systems 8](#_Toc396716514)

[2.3 Key Success Factors of Recommender Systems 8](#_Toc396716515)

[2.4 Recommender System Architecture 9](#_Toc396716516)

[2.4.1 Content-Based Recommender Systems 10](#_Toc396716517)

[2.4.2 Collaborative Recommender Systems 16](#_Toc396716518)

[2.4.3 Hybrid Recommender Systems 22](#_Toc396716519)

[2.5 Cold Starter Problem of a New User 23](#_Toc396716520)

[2.6 Change of User Preferences 25](#_Toc396716521)

[2.7 Suitability of Software Agents to Improve Scalability 26](#_Toc396716522)

[2.7.1 Multi Agent Systems 26](#_Toc396716523)

[2.7.2 Characterises of a Multi Agent System 26](#_Toc396716524)

[2.8 Review of Existing Systems 27](#_Toc396716525)

[2.9 Chapter Summary 29](#_Toc396716526)

[Chapter 3: Project Management 30](#_Toc396716527)

[3.1 Chapter Overview 30](#_Toc396716528)

[3.2 Project Management Methodology 30](#_Toc396716529)

[3.2.1 Time Allocation 30](#_Toc396716530)

[3.2.2 Constraints and Dependencies 30](#_Toc396716531)

[3.2.3 Potential Risks and Mitigation Plan 31](#_Toc396716532)

[3.3 Development Methodology 32](#_Toc396716533)

[3.4 Research Methodology 33](#_Toc396716534)

[3.5 Chapter Summary 33](#_Toc396716535)

[Chapter 4: Requirements Specification 34](#_Toc396716536)

[4.1 Chapter Overview 34](#_Toc396716537)

[4.2 Requirement Elicitation Process 34](#_Toc396716538)

[4.2.1 End User Questionnaire 36](#_Toc396716539)

[4.2.2 Formal Interviews 40](#_Toc396716540)

[4.2.3 Findings Derived From Requirement Elicitation 41](#_Toc396716541)

[4.3 Stakeholders 42](#_Toc396716542)

[4.3.1 Stakeholders and Roles 42](#_Toc396716543)

[4.3.2 The Context Diagram 44](#_Toc396716544)

[4.4 Use Case Diagram 45](#_Toc396716545)

[4.4.1 Use Case Descriptions 47](#_Toc396716546)

[4.5 Functional Requirements 49](#_Toc396716547)

[4.6 Non Functional Requirements 52](#_Toc396716548)

[4.7 Chapter Summary 52](#_Toc396716549)

[Chapter 5:System Architecture & Design 53](#_Toc396716550)

[5.1 Chapter Overview 53](#_Toc396716551)

[5.2 High Level Design 53](#_Toc396716552)

[5.2.1 Rich Picture of the NewsRec System 53](#_Toc396716553)

[5.2.2 High Level Architecture 54](#_Toc396716554)

[5.2.3 Article Descriptor and Reader Profile Modelling 55](#_Toc396716555)

[5.2.4 Recommender Algorithm Design 56](#_Toc396716556)

[5.3 System Design 56](#_Toc396716557)

[5.3.1 Selection of Design Methodology 56](#_Toc396716558)

[5.4 Domain Model of the NewsRec System 58](#_Toc396716559)

[5.5 Sequence Diagrams 59](#_Toc396716560)

[5.6 Design and Architecture Optimization 62](#_Toc396716561)

[5.7 Packages and Package Dependency Diagram 62](#_Toc396716562)

[5.8 ER Diagram 63](#_Toc396716563)

[5.9 Design Goals for Overall Solution 64](#_Toc396716564)

[5.10 Chapter Summary 64](#_Toc396716565)

[Chapter 6: Implementation 65](#_Toc396716566)

[6.1 Chapter Overview 66](#_Toc396716567)

[6.2 Technology Selections 66](#_Toc396716568)

[Selection of software agent development framework 66](#_Toc396716569)

[6.3 Implementation of REST Service 68](#_Toc396716570)

[6.3.1 Implementation of User Interaction Capturing Module 68](#_Toc396716571)

[6.3.2 Implementation of User Interaction Writing Module 68](#_Toc396716572)

[6.4 Core System Implementation 69](#_Toc396716573)

[6.4.1 Recommender Engine Implementation 70](#_Toc396716574)

[Chapter 7: Testing 63](#_Toc396716575)

[7.1 Chapter Overview 72](#_Toc396716576)

[7.2 Objectives and Goals of Testing 72](#_Toc396716577)

[7.3 Testing Criteria 72](#_Toc396716578)

[7.4 Functional Requirements Testing 72](#_Toc396716579)

[7.5 Module and Integration Testing 74](#_Toc396716580)

[7.6 Non Functional Requirements Testing 75](#_Toc396716581)

[7.6.1 Accuracy Testing 75](#_Toc396716582)

[7.6.2 Performance Testing 77](#_Toc396716583)

[7.6.3. Load and Scalability Testing 78](#_Toc396716584)

[7.7 Limitations of the Testing Process 78](#_Toc396716585)

[7.8 Chapter Summary 79](#_Toc396716586)

[Chapter 8: Evaluation 65](#_Toc396716587)

[8.1 Chapter Overview 80](#_Toc396716588)

[8.2 Evaluation Criteria 80](#_Toc396716589)

[8.3 Selection of Evaluators 80](#_Toc396716590)

[8.4 Evaluation Methodology and Approach 81](#_Toc396716591)

[8.5 Summary of Evaluation Survey Questions 81](#_Toc396716592)

[8.6 Evaluation Survey Findings 82](#_Toc396716593)

[8.6.1 Overall Concept 82](#_Toc396716602)

[8.6.2 Scope & Depth of the Project 83](#_Toc396716603)

[8.6.3 System Design, Architecture and Implementation 84](#_Toc396716604)

[8.6.4 Solution and Prototype 85](#_Toc396716605)

[8.6.5 Usability Accuracy and Performance 86](#_Toc396716606)

[8.6.6 Recommender Engine 86](#_Toc396716607)

[8.6.7 Limitations of the Solution and Future Enhancements 87](#_Toc396716608)

[8.7 Self Evaluation 87](#_Toc396716609)

[8.8 Chapter Summary 90](#_Toc396716610)

[Chapter 9: Conclusion 65](#_Toc396716611)

[9.1 Chapter Overview 91](#_Toc396716612)

[9.2 Achievement of Aim and Objectives 91](#_Toc396716613)

[9.3 Utilizing of Knowledge from Course Modules 92](#_Toc396716614)

[9.4 Use of Existing Skills 93](#_Toc396716615)

[9.5 Learning Outcomes 93](#_Toc396716616)

[9.6 Problems and Challenges Faced 93](#_Toc396716617)

[9.7 Limitations of the Research 94](#_Toc396716618)

[9.8 Future Enhancements 94](#_Toc396716619)

[9.9 Contribution 96](#_Toc396716620)

[9.10 Concluding Remarks 96](#_Toc396716621)

[REFERENCES i](#_Toc396716622)

[Appendix A – Gantt Chart a](#_Toc396716623)

[Appendix B – Gantt Chart b](#_Toc396716624)

[Appendix C – Test Cases g](#_Toc396716625)

# List of Figures

[Figure 1.1- Flowchart Describing the Basic Flow of the Proposed System 6](#_Toc396714481)

[Figure 2.1- Architecture Content-Based Recommender Systems 10](#_Toc396714482)

[Figure 2.2 - Architecture of Collaborative Recommender Systems 16](#_Toc396714483)

[Figure 4.1- Normal User Successful Response Rate 36](#_Toc396714484)

[Figure 4.2- Age Group Representation 37](#_Toc396714485)

[Figure 4.3 - News Site Access Frequency 37](#_Toc396714486)

[Figure 4.4 - Devices Used to Access News Sites 37](#_Toc396714487)

[Figure 4.5 - Smart Phone User Access Preference 38](#_Toc396714488)

[Figure 4.6 - Preference of News Article Access 38](#_Toc396714489)

[Figure 4.7 - Preferred News Article Finding Method 38](#_Toc396714490)

[Figure 4.8 - Preference for a Personalized News.Site.........................................................................................39 Figure 4.9 - Reasons for a Personalized News Site 39](#_Toc396714491)

[Figure 4.10 - Preferred News Categories 39](#_Toc396714492)

[Figure 4.11 - Reader Preference Indication 39](#_Toc396714493)

[Figure 4.12 - Non-Functional Requirement Rating 40](#_Toc396714494)

[Figure 4.13 - Onion Diagram for the Proposed System 42](#_Toc396714495)

[Figure 4.14- Context Diagram for NewsRec System 44](#_Toc396714496)

[Figure 4.15 - Use Case Diagram for System Login 45](#_Toc396714497)

[Figure 4.16 - Use Case Diagram for NewsRec System 46](#_Toc396714498)

[Figure 5.1- Rich Picture Diagram of the NewsRec System 53](#_Toc396714499)

[Figure 5.2- High Level Architecture Diagram 54](#_Toc396714500)

[Figure 5.3- Article Descriptor Model 56](#_Toc396714501)

[Figure 5.4- Reader Profile Model 56](#_Toc396714502)

[Figure 5.5- DFD Diagram of Interaction Capturing and Storing 57](#_Toc396714503)

[Figure 5.6- Domain Model of NewsRec System 58](#_Toc396714504)

[Figure 5.7- Sequence Diagram for New Reader Registration 59](#_Toc396714505)

[Figure 5.8- Sequence Diagram for Update Article Rate Value 60](#_Toc396714506)

[Figure 5.9- Sequence Diagram for Update Reader Profile 60](#_Toc396714507)

[Figure 5.10- Sequence Diagram for Recommendation of Articles 61](#_Toc396714508)

[Figure 5. 11- Packages and Package Dependency Diagram 62](#_Toc396714509)

[Figure 5.12- ER Diagram 63](#_Toc396714510)

[Figure 6.1- User Interaction Capturing REST Service 68](#_Toc396714511)

[Figure6.2- User Interaction Saving 69](#_Toc396714512)

[Figure 6.3- Invocation of Recommender Algorithms 70](#_Toc396714513)

[Figure 6.4- Collaborative Filtering 71](#_Toc396714514)

[Figure 6.5- Content Based Filtering 71](#_Toc396714515)

[Figure 6.6- Recommended News Articles 72](#_Toc396714516)

[Figure 7.1- Accuracy Rate of the Recommendations 75](#_Toc396714517)

[Figure 7.2- Accuracy Rate of the Recommender Approaches 76](#_Toc396714518)

[Figure 8.1- Non Functional Requirement Rating 86](#_Toc396714519)

# List of Tables

[Table1.1- Objectives 5](#_Toc396714612)

[Table 1.2 - Software and Hardware requirements 6](#_Toc396714613)

[Table 2.1 - An Overview of Recommender System Components and Their Roles 9](#_Toc396714614)

[Table 2.2- An Overview of Recommender Engine Components and Their Roles 9](#_Toc396714615)

[Table 2.3– Meaningful Representation of Item Descriptors & User Profile Information .................................10](#_Toc396714616)

[Table 2.4 - Overview of keyword based profiles 11](#_Toc396714617)

[Table 2.5- Overview of Semantic Network Profiles 11](#_Toc396714618)

[Table 2.6- Overview of Concept Profiles 11](#_Toc396714619)

[Table 2.7- Automated descriptor assignment approach 12](#_Toc396714620)

[Table2.8- Manual descriptor assignment approach 13](#_Toc396714621)

[Table 2.9 - Impact the Advantages of a Content Based Recommendation System 14](#_Toc396714622)

[Table 2.10 - Impact the Disadvantages of a Content Based Recommendation System 15](#_Toc396714623)

[Table 2.11- Overview of item based and user based approaches 17](#_Toc396714624)

[Table 2.12 - Impact of the Two Methods of Collaborative Recommendation System .....................................17](#_Toc396714625)

[Table2.13 - Impact on Accuracy by Clustering Algorithms 18](#_Toc396714626)

[Table 2.14- Impact on Scalability by Clustering Algorithms 19](#_Toc396714627)

[Table 2.15- Impact of the Advantages of a Collaborative Recommendation System 20](#_Toc396714628)

[Table 2.16- Impact of the Disadvantages of a Collaborative Recommendation System ..................................21](#_Toc396714629)

[Table 2.17- Negation of Disadvantages of a CBR system 23](#_Toc396714630)

[Table 2.18 - Negation of Disadvantages of a CF system 23](#_Toc396714631)

[Table 2.19- Implicit User Preference Capture 24](#_Toc396714632)

[Table 2.20 - Explicit User Preference Capture 24](#_Toc396714633)

[Table 2.21- Hybrid User Preference Capture 24](#_Toc396714634)

[Table 2.22 - Review of Existing System 28](#_Toc396714635)

[Table3.1 - Time Allocation 30](#_Toc396714636)

[Table 3.2 - Identified Risks for NewsRec System 31](#_Toc396714637)

[Table 3.3 - Characteristics of the Software Development Methodologies 32](#_Toc396714638)

[Table 4.1 - Evaluation of Requirement Elicitation Methods 35](#_Toc396714639)

[Table 4.2 - Limitations of the Questionnaire 36](#_Toc396714640)

[Table 4.3 - Limitations of the Interviews 40](#_Toc396714641)

[Table 4.4 - Summary of Findings 41](#_Toc396714642)

[Table 4.5 - Stakeholders and Roles 43](#_Toc396714643)

[Table 4.6- Use Case Description for Register to the System 47](#_Toc396714644)

[Table 4.7 - Use Case Description for Login to the System 48](#_Toc396714645)

[Table 4.8 - Function Requirements of NewsRec System 51](#_Toc396714646)

[Table 4.9 - Non Functional Requirements of NewsRec System 52](#_Toc396714647)

[Table 5.1-Overview of the Domain Model Classes 59](#_Toc396714648)

[Table 6.1-Comparisons of Agent Frameworks . 66](#_Toc396714649)

[Table 6.2- Agents Used in the System 70](#_Toc396714650)

[Table 7.1 - Tested Functional Requirements 73](#_Toc396714651)

[Table 7.2- Summary of Unit and Integration Testing 74](#_Toc396714652)

[Table 7.3 - Performance Testing of NewsRec System Using a Web Browser 77](#_Toc396714653)

[Table 8.1-Evaluation Criteria 80](#_Toc396714654)

[Table 8.2-Evaluator Groups 81](#_Toc396714655)

[Table 8.3-Evaluation Survey Questions 82](#_Toc396714656)

[Table 8.4-Summary of Evaluation Feedback on Project Concept 83](#_Toc396714657)

[Table 8.5-Summary of Evaluation Feedback on Project Scope and Depth 83](#_Toc396714658)

[Table 8.6-Summary of System Design, Architecture and Implementation 85](#_Toc396714659)

[Table 8.7-Summary of Solution and Prototype 85](#_Toc396714660)

[Table 8.8-Review on Evaluation of Recommender Engine Feedback 87](#_Toc396714661)

[Table 8.9-Self Evaluation of Evaluation Criteria 88](#_Toc396714662)

[Table 9.1- Contribution of Objectives Towards the Completion of the Project 92](#_Toc396714663)

[Table 9.2-Contribution of Modules Towards The Completion of the Project 93](#_Toc396714664)

[Table 9.3-Contribution of modules Towards The Completion of the project 96](#_Toc396714665)

# List of Abbreviations

|  |  |
| --- | --- |
| **Abbreviation** | **Definition** |
| **CF** | Content Filtering |
| **CBF** | Collaborative Filtering |
| **NLP** | Natural Language Processing |
| **OOADM** | Object-Oriented Analysis and Design Method |
| **SSADM** | Structured Systems Analysis and Design Method |
| **HTTP** | Hypertext Transfer Protocol |
| **SOAP** | Simple Object Access Protocol |
| **REST** | Representational State Transfer |
| **SRS** | Software Requirement Specification |
| **AI** | Artificial Intelligence |
| **RAM** | Random Access Memory |
| **JSON** | JavaScript Object Notation |
| **SOA** | Service Oriented Architecture |
| **JADE** | Java Agent Development |
| **UML** | Unified Modelling Language |

# 