

ARCHITECTURE OF THE SENTIMENT WEB MINING

Mohd Javed Mohd Nadeem Mansuri*

R.R. Keole**

Abstract

In this paper we had described the system architecture which is still in the development process whose purpose is to collect the sentiment of web users regarding various topic such as social issues like immigration, retail products, and financial instruments (FI).

The first step is acquiring of knowledge or knowledge acquisition. Various sources on the web helps a Sentiment Web Mining (SWM) system to acquire knowledge. Blogs, social networks, email, or online news proves the main place for finding such knowledge. A SWM system has personalization and customization capabilities. Customization occurs when the SWM user can change his/her preferences as it helps to select specific sites that can be used for data mining and evaluation. Based on the user profile, when the system decides which sites to be used for data mining, personalization occurs. The second step is storage of knowledge, which involves database creation. Indexing and tagging of appropriate web sites takes place. The hardest part of this step is Tax anatomy. A SWM system will use a series of off-the-shelf knowledge analysis/data mining tools including SWM knowledge analysis/data mining engine which is based on web services technology. The last step is widely spreading of knowledge to the users. The presentation component of a SWM system is separated from other components, namely, the process component, data access component, and business rule component so that they can be maintained easily.

* ME 1st Year (Computer Science and Engineering), H.V.P.M COET, SGB Amravati University

** Assistant Professor IT Dept, H.V.P.M COET, Amravati

1. INTRODUCTION

The word 'meme' as described by Richard Dawkins [6] as a cultural idea unit which gets transferred from one person to another by non-genetic means. Ideas can get spread quickly in social networks, emails, melodies, catch-phrases, images, and videos. Cool hunting is a term stated in the early 1990s which refers to a new race of marketing professionals, called cool hunters. Their job is to define changes that are predicted and do observations of new or existing cultural trends. The word derives from the aesthetic of "cool" [2, and 8]. There are several tools in the market to examine the social sentiment. Reference [9] mentions several social sentiment analysis tools.

Section 3 describes the requirement and overview of the SWM system that is under development. Section 9 describes the SWM architecture. The methodology used for SWM development is described in section 8 and 9. Section 4 describes the conclusion and future work.

2. Overview of the SWM System:

Requirements of the SWM system are captured through use cases [4] which are written in UML (Unified Modelling Language. For object-oriented analysis and design. IBM Rational Modeller [1] software is used. For development and deployment of the system Window 2008 R2 enterprise [7] is used which In turn uses several guest operation system and tools such as Window 2008 R2, Window 7, SQL Server 2008 R2. For development the Hyper-v [5] virtual machine mechanism is used. The virtual machine mechanism used for deployment is VMware [10].

SWM system consist of some issues and requirements which are stated below:

[1] SWM System is a web-based system and has a good user interface.

[2] SWM system has *personalization* capabilities. Based on the user profile when the system makes decision as to which sites should be used for analyses personalization occurs. There is a dynamic change in user profile which depends on the type of user request from

the system and the specific sites the user Visits to verify the result of the SWM system [11].

[3] SWM system can have the capability to both drill down and drill up.

[4] From email traffic [12] or social network web sites influential (leaders) gets identified.

[5] SWM system can support multiple languages.

[6] SWM provides trend analyses based on a timeline and a count of positive, negative or neutral

Public comments on a particular brand or subject.

[7] Using geo-tagging of websites, photographs, videos, or RSS (Really Simple Syndication), Identifies the geo-position of the entity (i.e.). For example the SWM should be able to produce a Sentiment for specific topic or product in the northeast, northwest, southeast or southwest.

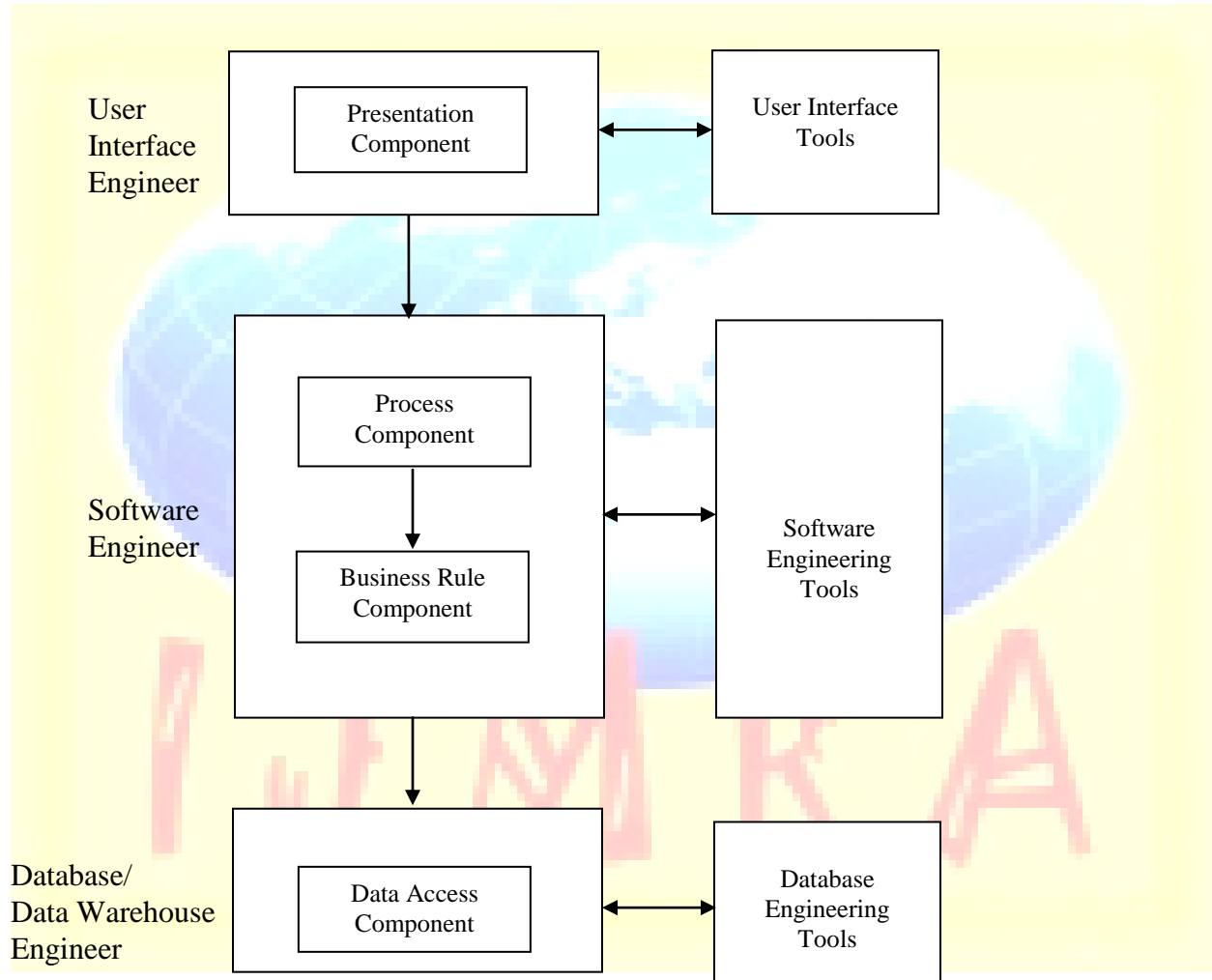
[8] The API (Application Program Interface) for an SMW system helps to get integrated With other application, especially applications that are dependent on data warehousing.

[9] SWM system has *customization* capabilities. When the SWM user has the capability to change his/her preferences (Web sites, positive, negative verbs, or phrases) so as to select specific sites that can be used for different analyses, customization occurs. Based On a specific web site, individual or chatting user name, the user can also rate the comments (intensity) .

[10] SWM system is designed so that it has the capability to integrate with other semantics analysis and data warehouse data mining tools having an open architecture.

3. Social Network Bus Architecture:

In order to develop a multi-tier architecture for a SWM system Concurrent engineering methodology is used [13]. Integrating different disciplines such as software, user interface design, data warehouse engineering, and database from the beginning of the project is required for Concurrent engineering in software development. Below is the Figure of Components of Software System.



There are basic four components that make up the Software systems architecture: 1) process component defines which software component can be processed by the business rule component. A software engineer works on this component. 2) Processing of the business rules of the organization which are dictated by a process component is done by business rule component. Main working on this is done by Software Engineers since business rule are mostly

dynamic in nature and a rule engine implements them. 3) The actions such as insertion, updation, deletion, and retrieving from the databases &/or data warehouse are done by the data access component. This layer holds all the commands relating to SQL (Structure Query Language) only when a relational database is chosen for all the system. This components makes use of database and data warehouse engineers. 4) Presentation component which is also a system interface which has users such comprising of web, graphical user interface, or voice response unit. This component makes use of user interface engineer and also the Database and data warehouse engineer works on this component. Database and Data warehouse Engineer engineers also play an important role for the logical as well as physical design of a database and/or data warehouse. In order to avoid redundancy normalization techniques (usually a third normal form) is used. A dimensional model is used in a data warehouse that allows redundant data. Figure 1 shows typical software components [3].

SWM system provides the social network bus architecture for creating standards for different components of social network and third party integration. The bitmap is used for financial instrument taxonomy. The financial instrument bond maybe considered as fixed income for a group of user and mutual fund for another set of users depending on interpretation of financial instrument bitmap characteristic by individual user. In other words SWM system is flexible to allow the user to define what their interpretation of financial instrument is.

4. Conclusion:

In this paper, what the Sentiment Web Mining (SWM) system requires is defined. A series of sentiment analysis components gets integrated in a web based SWM system. It has many features including open architecture and also other properties such as volume (positive, negative, and neutral public opinion), intensity, geo-position, and trend analysis of social network web sites, emails, images and videos. Influential (leaders) from email traffic or social network web sites are identified by the SWM System.

The SWM system gets developed using Object-oriented and concurrent engineering methodology. With clear separation of presentation, process, business rule, and data access component, SWM system is multi-tier architecture. With other software components it also provides API, web services, and XML interface for ease of integration. For financial

instruments which uses a bitmap characteristic, special taxonomy is also used. The users gets a very useful hand on interpretation of complex financial instrument by using this taxonomy.

5. Future Research:

Below the Future Research and work is listed:

- [1] Special domain such as retail, politics, and financial instruments gets more taxonomy algorithms developed for them.
- [2] In order to have a more advance user interface and reports, especially for financial instruments development and research is done.
- [3] Using massive parallel processors more advance natural language processor are identified or developed.
- [4] Using associative memory structure and unstructured data is faster retrieved.
- [5] Correct results of SWM System and other sentiment analysis tools are estimated with respect to their percentage.

6. References:

- [1] <http://www-01.ibm.com/software/awdtools/modeler>
- [2] <http://en.wikipedia.org/wiki/Coolhunting> - Coolhunting
- [3] Hyper-v Server 2008 R2 - <http://www.microsoft.com/hyper-v-server/en/us/default.aspx>
- [9]
- [4] VMware - <http://www.vmware.com/solutions/cloud-computing>.
- [5] Social Sentiment Analysis Tools
 - a. www.lexalytics.com
 - b. www.twitrratr.com
 - c. www.sentimentmetrics.com

- d. www.netbase.com
- e. www.searchenginejournal.com/alertrank
- [6] Dawkins, D., Oxford University Press, 1976, Oxford, New York. The Selfish Gene, ISBN-13: 978-0192860927.
- [7] Monimoto, R., Noel, M., Droubi, O., Misty, R., Amaris C., Sams Publications 2010. Windows Server 2008 R2 Unleashed, ISBN-13: 978-0-672-33092-6.
- [8] Swain, G., Laxmi Publications, 2010. Object-Oriented Analysis and Design Through Unified Modeling Language, ISBN-13: 978-9-3803-8654-6.
- [9] Movafaghi, S., Chan, T., Pournaghshband, H., Collins, S. The Customization and Personalization for Library Portal. The Fifth International Conference on the Book. October 20-22, 2007.
- [10] Gloor, P.; Dynes, S.; Zhao, Y. Visualization of Communication Patterns in Collaborative Innovation Networks - Analysis of Some W3C Working Groups. Proceedings of the twelfth international Conference on Information and knowledge management, New Orleans, LA, USA, Pages: 56 - 60, 2003, ISBN:1-58113-723-0.
- [11] Movafaghi, S.; Pournaghshband, H.; Collins, S. Concurrent Engineering Measurement: Database and Data Warehouse Engineer. The 2007 International Conference on Software Engineering Research and Practice (SERP'07: June 25-28, 2007)
- [12] Movafaghi, S.; Pournaghshband, H.; Collins, S. Concurrent Engineering: A Roadmap to Software Engineering, Database and Data Warehouse Technology. The 2007 IRMA International Conference.
- [13] Gloor, P. Cooper, S. 2007. Coolhunting - - Chasing Down the Next Big Thing? AMACOM, NY, ISBN-13: 978-0-8144-7386-3.