

# Yammer: A Study on Knowledge Sharing using Enterprises Microblogging System

Angela Lee Siew Hoong<sup>+</sup>, Tong Ming Lim and Justin Lim Renn Aun

Sunway University, 5, Jalan University, Bandar Sunway, Selangor, 46150 Malaysia

**Abstract.** Understanding text data on social networking systems has become an important task for companies nowadays. The shift from pattern mining of structured database to mining of non-structured text data has awoken company operators to have a strong presence in the world of social media. This research uses SAS text miner to analyze conversational text collected from an enterprise microblogging system, Yammer. All posts on Yammer are called “Yams” (Kelsey Blair, 2011). These texts are used to study the topics of discussion among employees and the knowledge sharing activity of employees in the case company by analyzing the topics maps. The aims of this research would be on the topics of discussion and the frequency of topics discussed among employees that can be used to determine whether the microblogging system is a potential platform to facilitate better knowledge sharing and knowledge creation among employees. The case study company in this research project is a knowledge centric organization involves in some forms of knowledge sharing activity using a knowledge management system. The methods used in this research would be choosing members that are involved in an active project to use Yammer instead of the current KMS system to study the frequency of communication and topics of discussion from their conversational text. The text analysis diagram and topic map are generated in order to identify the patterns of discussion among employees and the topics exchanged between employees on the microblogging platform. SAS (Statistical Analysis System) text mining tool was used to conduct the text mining analysis where a number of visual presentations were developed to study the communication patterns among employees. The results of this research had shown that text mining is able to surface employees’ frequency of communication and topics of discussion through posting activities and the informal knowledge exchange activity among employees is a form of knowledge creation and sharing which can be used as the future research works.

**Keywords:** Text mining, enterprise 2.0, microblogging, usage pattern, communication behavior

## 1. Introduction

Hidden knowledge in social media systems such as Facebook and Twitter has become an important source of knowledge for companies and nonprofit organizations. Plenty of research works conducted in the last few years have shown that design of new products and launching of marketing campaign were crafted based on feedbacks and comments mined from unstructured data posted by customers on these platforms. Understanding customers’ needs and critiques before investing into commercial activities save large amount of money and the outcomes harvest from the investment is more encouraging.

However, mining text information is not an easy task. Text data is ill-structured and hence proper tools are required. In addition, the amount of data produced by the bloggers and other content producers are usually dynamic and this makes text mining a challenging task indeed. In this research, an enterprise microblogging system, Yammer, is chosen and implemented for shared services companies. An enterprise microblogging system runs behind the firewall. Compare to Twitter, which is a public microblogging system, expose the ‘twitts’ to their followers openly. An enterprise microblogging system, on the other hand, requires that the participant belong to a specific company domain. Linkage to the external companies is done through an external community linked. This is accomplished by invitation only. This research was conducted to identify the frequency of discussion and topics discussed from the posts which were entered on the enterprise microblogging system by the employees of a case study company. Unlike data mining which mines structured data in database systems such as Oracle or MS-SQL, which serves as the container for the knowledge repository, text mining focuses on mining unstructured data such as text which are then converted into meaningful knowledge. Data that contain simple “yes” or “no” answers are easily analyzed; however,

---

<sup>+</sup> Angela Lee Siew Hoong . Tel.: + 0163522281  
E-mail address: [angelal@sunway.edu.my](mailto:angelal@sunway.edu.my).

extracting knowledge from the text is a challenging issue. This is due to accuracy of the data that the users required is not easy to obtainable. The detailed analysis of text requires thorough understanding of the domain to make sense. Therefore, the use of text mining software such as the SAS Enterprise Miner (E-Miner) that has the capability to analyze text data in order to convert it into useful statistical and visual representation is an essential to this research.

## **2. Related works**

Ning Zhong et al (2010) adopted Natural Language Processing (NLP) technique, a modern computational technology approach, to mine the meaning behind the text documents so that they are easy to understand. This approach allows the text mining engine to discover relationships or patterns that exist between texts in the files. Moria Levy (2009) reckoned that younger employees in organizations can serve as knowledge catalysts to start using WIKI or blogs. She pointed out that WEB 2.0 concepts should be tested as to organization's maturity to decide if they can be adopted as part of the organizational knowledge sharing. Her paper also highlight that better assimilation of knowledge management may be triggered by the WEB 2.0 phenomenon. The Web 2.0 concept (Chatti et al, 2007) is also recognized as an ideal fit with Nonaka's SECI approach in which it opens new doors for more personal, dynamic, and social learning on a global scale. Web 2.0 presented an extended view of blended learning which includes the combination of formal and informal learning, knowledge management into one integrated solution, by discussing what we call the Web 2.0 driven SECI model based learning process. Maria R. Lee et al (2007) pointed out that the conventional centralized knowledge management system has been used to handle large volumes of information since the instigation of World Wide Web. However, the authors highlighted that knowledge residing in the repositories has not been accumulated or integrated to generate new intelligence. New communication technologies have changed the paradigm in knowledge management. They analyzed the recent web trends on collaborative intelligence applications that support for the change which provides the opportunity for more effective and feasible knowledge management. The impact of Web 2.0 on knowledge management and the new orientation of KM (Shimazu et al, 2007) to organizations with user participation-type culture and the "collective intelligence" approach have attracted attention of company operators. Opinions are currently being advanced with regard to the concept and process of KM and the system architecture that can be used to implement it. The authors introduce a KM model in the context of the Web 2.0 that can expand collective intelligence in a positive spiral by closely linking it to knowledge extraction from various communication tools and job systems.

## **3. Research Methodology**

The research work that was conducted adopted the following methodology. First of all, an active project that consists of members from all the departments that participated in the project was identified. This was followed by a briefing that set the objectives of the test clearly. The duration of the test was also agreed and identified. Yammer was chosen as the enterprise microblogging tool that will be implemented and tested by the chosen company for a period of 30 days by members of a project team. Training on Yammer is provided. All the entries posted by participants were extracted into an Excel file. The Excel file contains the raw post data. The raw post data is imported into the SAS text mining system so that the raw post data could be converted into SAS file format. In this research project, the SAS Text Miner is used to analyze the microblog entries so that unstructured text that consists of the employees' opinions could be identified and analyzed. Since the data contain more than just simple "yes" or "no" answers or plain numerical values, the need to process and analyze such complex unstructured inputs is necessary in order to provide human understandable outputs. By analyzing these entries, the text mining software is able to identify similarities and differences from these texts. Hence, once the conversion of input texts have completed, these files will be used by the software to mine to produce frequency diagrams and topic maps for analysis purposes. The SAS files that contain all the data extracted from the Yammer software is the data source for the analysis work. The analysis is able to identify the patterns of the conversation among the employees of the selected project. These processed outputs provide a better and accurate understanding of the patterns for the chosen

microblogging system and usage pattern of the employees. The analysis from the easy-to-understand visual presentation will highlight the concept of interests that most employees post on Yammer.

#### 4. Empirical Analysis

The SAS Enterprise Miner software is used to construct a process diagram (Figure 1). The text miner node connects to the data source which instructs the text miner to process the input text to provide visual presentations such as concept maps. Figure 2 depicts the patterns of the post entries of the employees visually on their usage in the form of frequency chart by author. Figure 2 shows that the frequency of entries posted by different authors. It is shown that one of the project members, Phoebe Than is the most active participant among all the other participants in the microblogging space. She has posted the most entries in this exercise. She had contributed a total of 61 entries. This may be due to her position as project manager of the chosen project. Her support on the use of microblogging system to update status and share knowledge with her other peers improve sharing activity on Yammer. Her contribution as a knowledge creator and sharer has also ensured that the research project yielded the anticipated outcome. Based on the result as illustrated in Figure 3, it has clearly shown that the frequency of terms and the weight of each term highlight the communication pattern and the importance of topic discussed among employees. For example, the box which is highlighted in red shows that the term “+server” has the highest frequency count. In this example, it has a total of 52 counts that appear in the data collected from the entire test period. However, the weight of the term is only 0.56547 which shows that the importance of this term is at an average level. The higher the weight is the more importance the term is from the entire corpus of the text compiled in this experiment. Therefore, this has shown that the entries posted in the project are mostly related to the “server” issues. Question that one may need to answer in the next phase of the research work is to understand the low weightage value for a term that was identified as a slightly above average term from the source input.



Figure 1: Flow of Text Mining

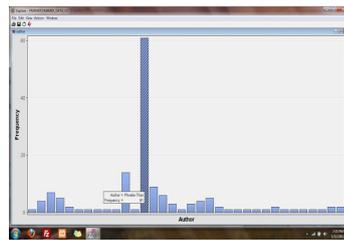


Figure 2: Author with Postings Frequency

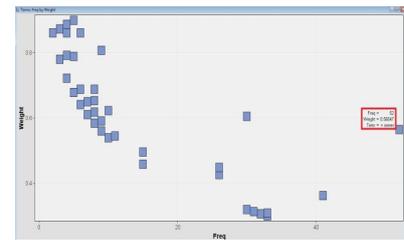


Figure 3: Terms with Highest Frequency

In Figure 4, it shows that the term, “license” was found to have the highest weight with a value of 0.9. This means that the importance of this term is the highest among all other terms being used. Although the frequency of this term is not very high, it has only five (5) occurrences in the entries posted by the users. However, it is an important term in this project. This pattern conveys a message that the current project is handling the “license” or legal issues. This shows that the importance of this term is relevant to the project. In Figure 5, a concepts map is used and presented to show the topics which are related to the term “ITSSC”. Some of the terms which are believed to be useful could be expanded into a more detail level by drilling into the second level of the map to understand the topics of communication. For example, from the concepts map in Figure 5, the terms “setup” is expanded and it shows that “setup” is relevant to the infrastructure department and it is also related to the setup of “server” term. Furthermore, the term “Putra” when it is expanded shows that it is related to the term “Sunway”, “Project” and “Hotel”. This explains that several topic of discussion among participants are all related somehow to the Putra Place project. Furthermore the lines that connect to the terms were presented using different thickness. The lines connecting the term “Putra” and “Hotel” is the thickest which means that the frequency of these two terms is the highest. With regard to the term “Hotel”, it is shown that the term is related to the “Pyramid”, “email domain” and “implementation”. These connections are relevant because our interview data toward the end of the testing period has verified that the finding of the patterns of the software has matched the interview data collected. While trying to identify the terms used from the content of the microblogging system, the text analysis software also find another term which was high relevant to the project.

The term “legal privileges” and all other relevant concepts are illustrated in Figure 6. Since the project is now discussing the legal issues with its clients, therefore conversation between the members are also mostly focus on the legal precaution that is of high concerns among member in the project. Hence, it can be seen in the concept map that the patterns of this topic and other topics that are very related to the legal matters such as “unauthorized”, “+prohibit”, “liability”, “virus damage” and “legal” terms. Furthermore, the thickness of the line connecting each terms are almost the same which shows that the frequency of these terms appearing together are almost the same. In addition, the term “legal” was expanded to the second level which visually explains other related terms such as “consent”, “companies”, “disclosure” and “responsible” are also relevant. The concept maps reveals the conversation patterns on the subject “legal privileges” are mostly related to each other and is relevant to the conversation of the project members. The concept map in Figure 7 shows the topic pattern of the term “Putra” that comes from the Putra Place project are related to the term “Putra” when expanded to the second level shows how other terms are related to each other. The concept map in Figure 7 validates that the term “Putra” is closely related to the term “hotel” with the line thickness that explain the relationship of these terms each other. Furthermore, based on the entries in result analyzed by the text miner software, the second thickest related term is “place”. This shows that the most related entries are about the network and the knowledge of network from the terms “join” and “knowledge” as clearly indicated in the concept map in Figure 7.

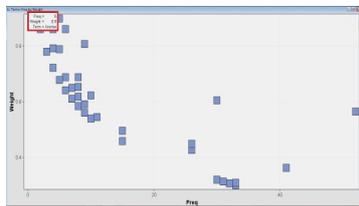


Figure 4: Terms with Highest Weight

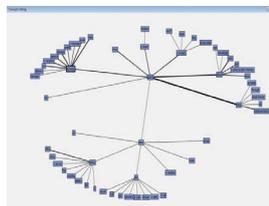


Figure 5: Concept Map for "ITSSC"

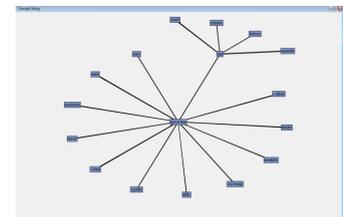


Figure 6: Concept Map for "Legal Privileges"

This pattern is related to the entries being posted in Yammer from a close raw text data extracted from the conversation among participants. As shown in the concept map, the term “team” is related to “sales”, “marketing”, “finance”, and “purchasing”. It was found that new groups were created by the users in this exercise. This shows that users actually make use of Yammer to create new CoP for their own departments. The text parsing and text filter analysis was also conducted in order to extract information on specific terms used by the participants. Figure 8 shows the updated activities sequence diagram. Based on the updated activity diagram, new findings for the information retrieval are presented (Figure 9). By using the Text Filter node, information can be easily retrieved. The Interactive Filter Viewer shows the communicated topics of related the authors visually. In Figure 9, when terms such as “implementation” and “server” are entered as the retrieving criteria, the documents that contain these terms with author detail are presented visually. Furthermore, the software also shows the Relevance column which indicates how relevant the term is.

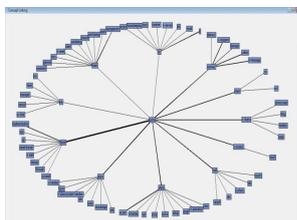


Figure 7: Concept Map on the Term "Putra"



Figure 8: Flow of Information Retrieval

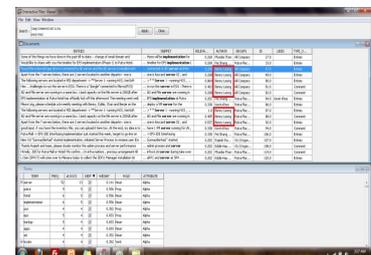


Figure 9: Retrieving Author Based on Terms

## 5. Analysis and Evaluation of Results

Based on the findings, the most and least communicated topics were illustrated through the use of diagrams and concept maps. The patterns found have helped to identify the communicated topics among employees. The posting activities and the content of the posts reveal that Phoebe Than contributed a fair amount of entries. She has taken the knowledge contribution champion role to create and share knowledge

with members in the team. Since the project is almost at its completion stage, the most discussed topic was on the legal issues. The patterns identified from the posts on the term “license” carries the highest weight and the term “+server” carries the highest frequency matched the content of the post-analysis interviews. The reliability of the result is high because licensing issue is the current topic of discussion. Furthermore, based on the observations carried out on the post activity, the term “+server” was the most talked about topic among members. On the pattern analysis, the topics of discussion were shown in the concept maps which clearly display the knowledge exchange pattern among employees. For instances, the concept map on the term “legal privileges” is connected to “unauthorized” and “legal”. This shows that these topics discussed are related and relevant to the project do take place. Furthermore, the text filtering technique allows information retrieval through the use of high frequency terms. By entering one or more terms, the documents that have these terms will be retrieved. It is clear that the text filtering can be used to identify authors and their entries which are related to the chosen terms. In a nutshell, the patterns of the analysis show that the project members make use of Yammer to share knowledge on the progress status of the project. Although the testing period was short and the number of post entries is small, the findings show that entries posted are project related. Therefore, microblogging would be useful to be used within companies for knowledge sharing if it is given sufficient resources such as time and larger group of participants actively involved in this exercise.

## 6. Conclusion and Future Research

In short, the findings of the research have shown that the communication patterns and topics of discussion among employees based on the “Yams” on the Yammer platform have signs of improvement compare to the existing KMS. Therefore, microblogging is a potential tool to share knowledge. The sharing behavior and adoption of Web 2.0 tools on employee productivity and job performance in companies pave a strong foundation for future research works.

## 7. References

- [1] Ning Zhong, Yuefeng Li & Sheng-Tang Wu, (2010), *Effective Pattern Discovery for Text Mining*
- [2] StatSoft Electronic Statistics Textbook, (2011), *Text Mining*
- [3] Maria R. Lee and Yi-chen Lan (2007) "*From Web 2.0 to Conversational Knowledge Management: Towards Collaborative Intelligence*", Journal of Entrepreneurship Research June 2007, Vol.2, No.2, p. 47-62
- [4] SHIMAZU Hideo, KOIKE Shinichi (2007) "*KM2.0: Business Knowledge Sharing in the Web 2.0 Age*" NEC TECHNICAL JOURNAL Vol.2 No.2/2007 Pp. 50-54
- [5] Kelsey Blair (2011) "*How to Use Yammer... And Why You Should Be Using It At Your Business*"
- [6] Chatti, Ralf Klamma, Matthias Jarke, Ambjörn Naevé (2007) "*The Web 2.0 Driven SECI Model Based Learning Process*" Advanced Learning Technologies Seventh IEEE International Conferences Pp. 780-782
- [7] Moria Levy (2009) "*WEB 2.0 implications on knowledge management*" JOURNAL OF KNOWLEDGE MANAGEMENT VOL. 13 NO. 1 2009, pp. 120-134