

Text Mining Analysis on Knowledge Sharing Using Enterprise Microblogging System

Angela Lee Siew Hoong,
Prof Lim Tong Ming,
Justin Lim



INTRODUCTION

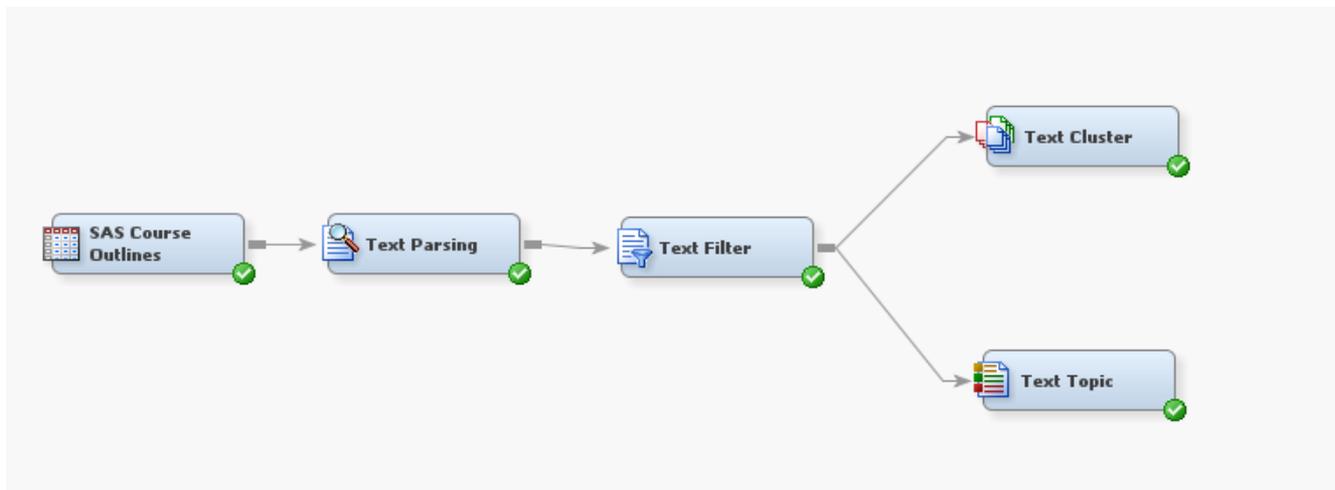
- Facebook and Twitter are social media systems that hold hidden knowledge which may be a source of information for companies.
- It is crucial to understand customers' comments before investing into any commercial activities. As a result, lot of research works were conducted based on feedbacks and comments mined from the unstructured data posted by customers on microblogging platforms before any important decision is made.
- However, mining text information is not an easy task. One needs proper tools to mine these data efficiently.
- Amount of data produced by the bloggers are usually dynamic and this makes text mining a challenging task indeed.
- In this research, Yammer, an enterprise microblogging system, is implemented for a company.

Research Goals

- The goal of this research is to study knowledge creation and sharing activities in the company. The factors of consideration on the choice of company are based on the following criteria:
- The company has started practicing knowledge activity using some knowledge management systems.
- The company sees value or needs where knowledge make competitively different in a globally open market.
- The company also sees organizational knowledge as an asset for decision making, improve productivity, performance and quality from the perspectives of company direction by the management.

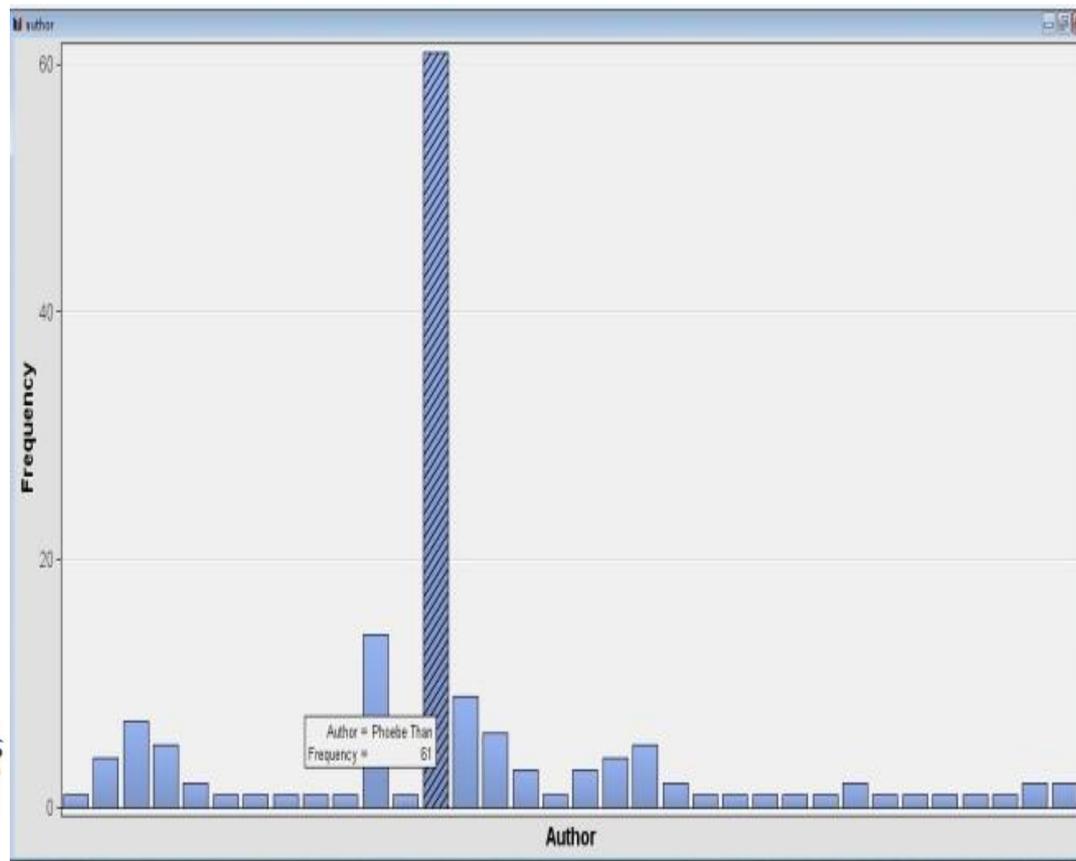
TEXT MINING USING SAS

- The analysis carried out in the research used a chain of process nodes to construct a series of activity nodes to analyze the raw text and generate analytic outputs to be interpreted.
- These nodes are used to connect to the data source which instructs the text miner to analyze the input data to provide text and visual presentations.



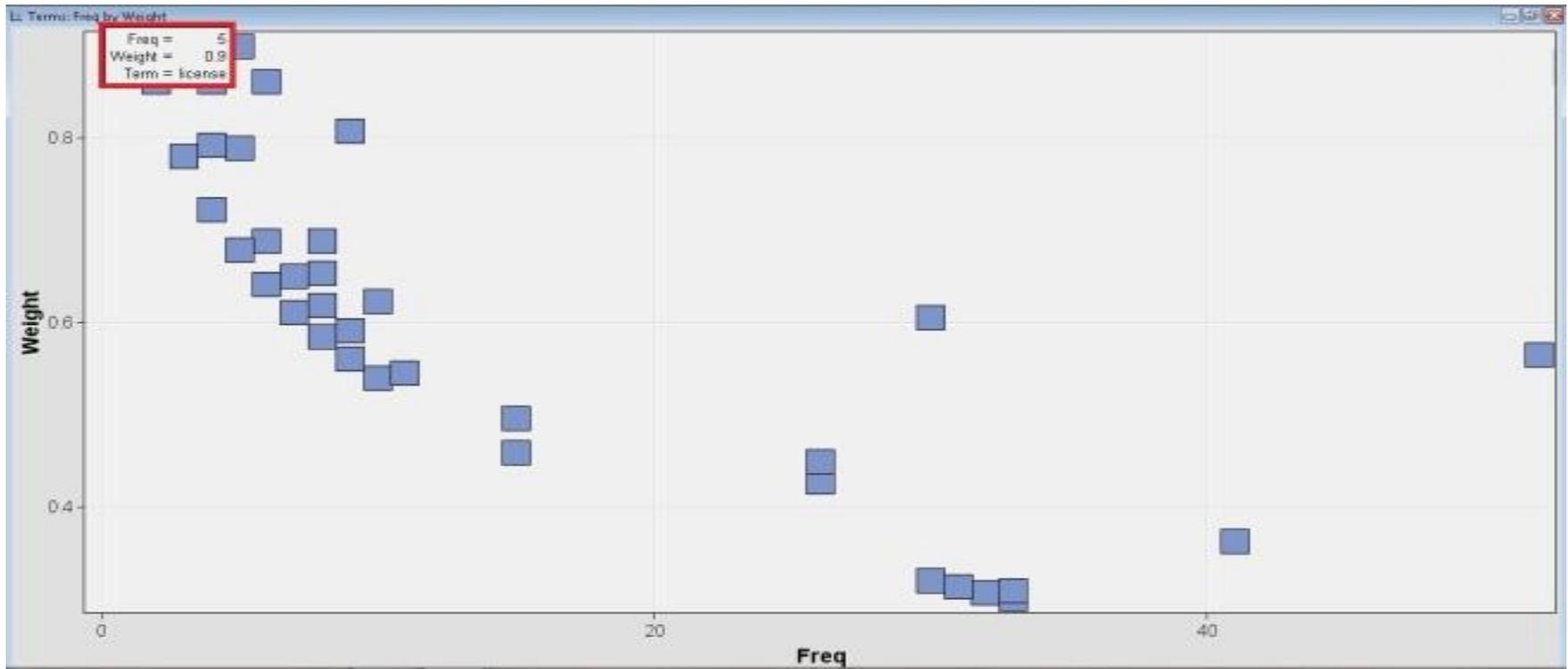
TEXT ANALYSIS – TEXT FILTER

- Depicts the communication and interaction patterns of the post entries
- Their usage is presented as visual outputs. Their usage is shown in the form of bar chart indicating frequency of post by author.



TEXT ANALYSIS – Terms by weight

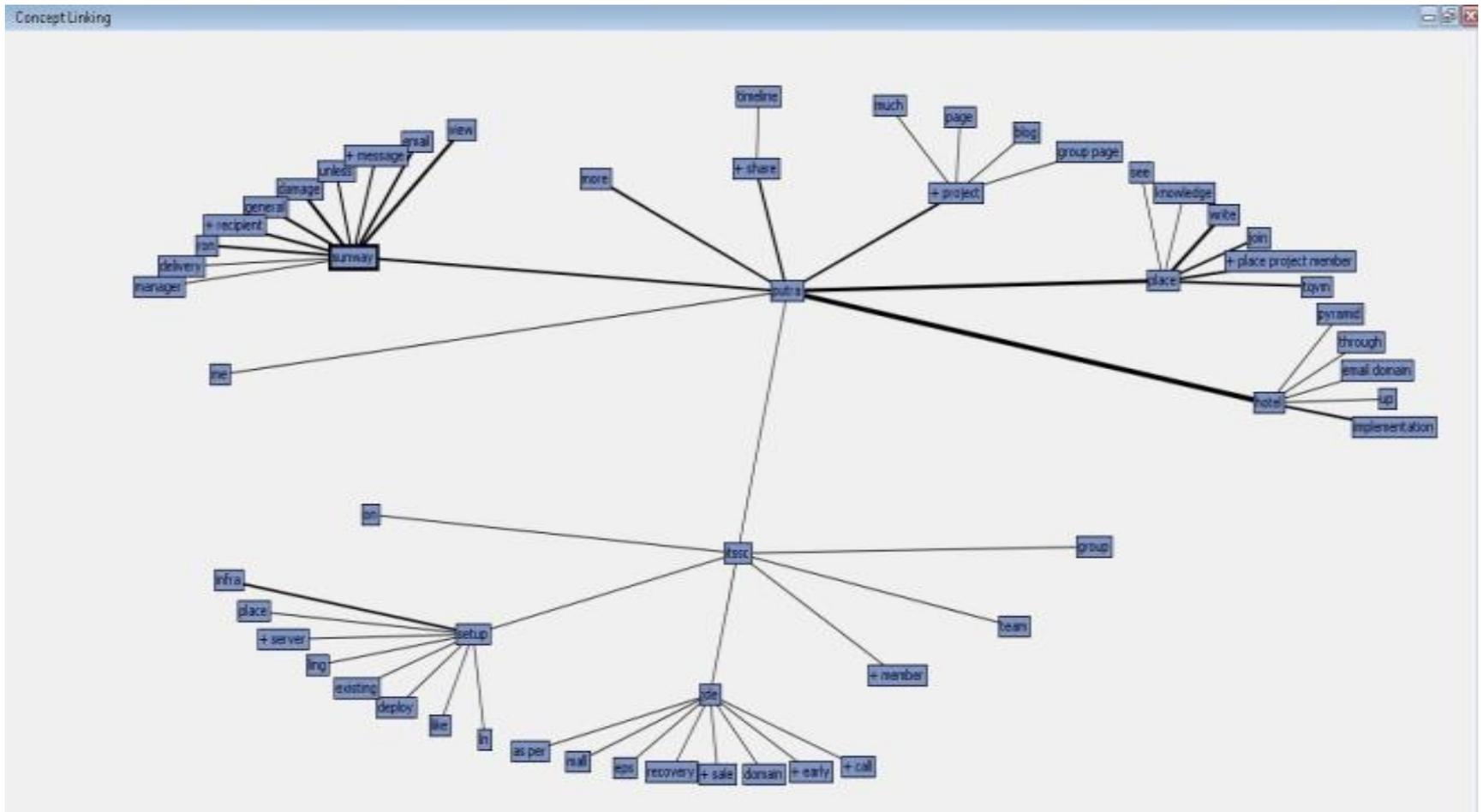
- In this figure it shows that the term, “license” had a value of 0.9 which was found to have the highest weight. This simply means that the importance of this term is the highest among all other terms being used and the frequency of this term is not very high, it has only five (5) occurrences in the entries posted by the



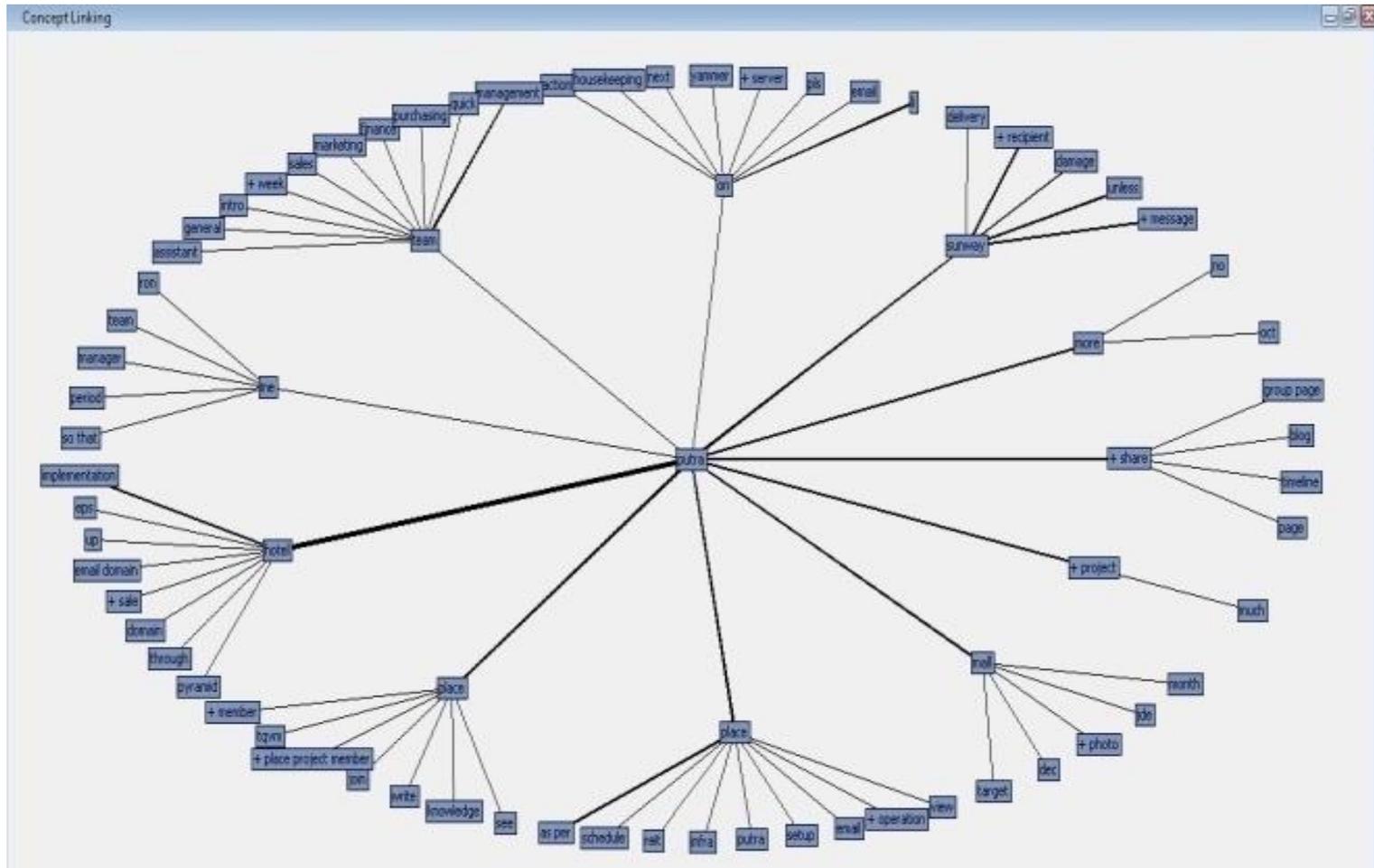
TEXT ANALYSIS – Concept map

- Concept Map is used to present the discussion topics that are related to the term.
- It discussed among employees through their weights.
- The thicknesses of the lines that connect these terms are an indication of the importance of the concept discussed among employees. The lines connecting the term “Putra” and “Hotel” is the thickest which means that the frequency of these two terms is the highest.

TEXT ANALYSIS – Concept Map



TEXT ANALYSIS – Concept Map



RESULT DISCUSSION

- The most and least communicated topics were illustrated through the use of concept map or concept linking diagrams.
- The use of topic map is to illustrate the topics, occurrences and association of topics discussed among employees whereas social network graph illustrates actors or ties in a network, relationships, connections and interactions of these actors in the network.
- The patterns from the analysis have helped the analyst to discover and identify the communication topics among employees.

RESULT DISCUSSION

- 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.
- The text filtering technique allows information retrieval through the use of high frequency terms.

CONCLUSION

- The findings have shown that the communication patterns and topics of discussion among employees on the enterprise microblogging system have shown signs of improvement compare to the existing classical KMS.
- The terms that have the highest occurrences and weight are the most important topics that are frequently talked.
- Some topics were selected to illustrate the significant increase in the communication frequency between employees. This shows that text mining is able to discover discussion patterns, terms and concepts shared among employees.