Microblogging and Data Mining for Knowledge Management

Part One

Benefits of Using Microblogging for KM

Knowledge is defined by McDermott (2000) as an output derived from social construction. With that being said blogging is powerful in spreading information, solicit comments and links and categorising entries. As one of the knowledge management tools blog is an advanced platform for effective information and knowledge sharing when compared to other traditional methods such as e-mails and discussion forums (Chan, Chu, Lee, Chan and Leung, 2013). Many researchers have found blogs to be very effective in developing and maintaining community relationships.

Knowledge sharing has been perceived as being very time consuming before, but since the microblog posts have a limited length they have the power to minimise information overload and attract increase participation compared with other Web 2.0 applications (Stocker, Richter and Riemer, 2012). An example of this is Twitter, instead of writing one lengthy post on one topic a day, through the limited length of characters, one is encouraged to make posts more often on different topics each day.

Many organisations have a similar goal which is to reduce their amounts of daily emails. The introduction of micro-blogging in enterprises and other social media application may be a solution. At Aton Origin managers are estimated to spend more than 5 but less than 20 hours per week for reading and writing emails. This has a very detrimental effect on production. This means that the amount of time spent on emails is more than a two day shift if each shift is equals to or less than 8 hours. If a manager can have more time by eliminating the time taken on emails he/she can increase his/her productivity. As two days are wasted on emails than this means the manager only has three days to do his other job per week if he works five days a week.

A response by one of the employees at Siemens that was asked to comment about micrologging that is used in the company said “it helps us understand and become aware of the latest news in terms of product releases, features and market enhancements. It is also important to understand the on-going challenges, projects delivered and business challenges in other business sectors”.

Problems of Using Microblogging for KM

Even though the limited character length in micro-blogging reduces information overload it also comes with limitations that hinders its success. When the message cannot fit in the allotted space micro-bloggers tend to abbreviate their words or rather change the word spelling in an effort to reduce the number of characters to meet the limited character length (Carter, Weerkamp and Tsagkias, 2012). This gives rise to text messages that are very difficult to comprehend especially if you are new to the environment. A word can be abbreviated or its spelling can be changed in many
ways depending on the person writing it. In other words there are so many casual ways of re-writing them. For example one can re-write the word ‘writing’ in these ways: “riting”, “ryting”, “ritin”, “rytin”, “yting”. Or the word ‘great’ can be re-written in these ways: “gryt”, “gr8”, “gret”. All these re-written words have lesser character number than in their formal spelling. Even though they are short they may be confusing to another person who is a recipient and is needed to respond. The message might be useless to him then.

Another effect of shortened words is that it makes it hard to identify language used by the sender when writing the message. When some words are changed they appear in another language but you may find that the language interpreted by the recipient is not the one used by the sender. For example in the shortened text above, is this English or Afrikaans? One might struggle to differentiate between the languages.

In an effort of trying to combat information overload through the limited character length in micro-blogging than a company can create its own micro-blog language which consist of all the abbreviated or shortened words. This language can be standardised in micro-blogging to minimise confusion in shortened words. If a company does not do this than it is hindering the success of the micro-blogs.

Limited character length that is used to eliminate information overload also influences it in return. In other word it causes or creates two relationships to associate with information overload, one that eliminates it as described above and the one that influences it. As there is an allotted character space a person can utilise to send one post, if he or she feels that content posted on the first post is not sufficient to get the message across he or she may make further posts until he is sure that his or her message is clear. The initial reason created to eliminate information overload now influences it. This is like a snake eating its tail.

References


Part Two

Effectiveness of Mining a Microblog

When talking about mining an ordinary human being would thing of situation whereby someone is digging the ground and going under it in search of valuable minerals such as gold, coal or diamond etc. but because of innovations that is not how it should always be as the term ‘mining’ may be used for differently. When used in IT it is called Data Mining. This is a process of digging through and analysing an enormous amount of data and then extracting its meaning (Alexander, 2014). As in mining gold there is machinery that is used to do such, also in data mining there are tools that are used. These tools can answer questions that were generally time consuming to solve.

The rationale driving the social network information is that it can disclose underlying motivation to create and share content while providing an elegant way to reorganize topically diverse tweets. This makes topic mining in microblogs much simpler. Tweets are particularly customer or user based as they come directly from the user’s own interests. They are not a generalization or do not represent a sample of a population or something else. When studying them you do not focus on the group or the large part to understand the individual members but you focus on individuals to get an understanding about the group. Moreover there exists tweets in the form of conversations and re-tweet messages that consist of the information that is structured on the social networks (Liu, Yin, Ouyang, Huang and Lin, 2013).

The human annotated tags semantic tags in tweets and hashtags are a useful resource that upgrades the topic mining part of data mining. For example some of the tweets that the Information Systems and Technology honours students in UKZN posted in Twitter had the #UKZN2014 hashtag in them. When one needs to mine this microblog data for a certain use he does not need to now names of the students that are posted and search individually their tweets that they made. But he can use #UKZN2014 as a search keyword to retrieve all the tweets that were made by UKZN students in 2014. This feature of a microblog search saves a lot of time for the data miner as there are so many tweets that are made in Twitter each and every day, estimated at about over 340 million in 2012. The ability to eliminate all the unwanted tweets over so many is something worth talking about. It is important also to note that these results of UKZN students are for 2014. On all the tweets that were made by UKZN students it retrieves those for this year. If the data miner were to use other hashtags like #usefulinfo it will bring back tweets that were made by people around the world that assisted with useful information. Now trying to find UKZN student’s tweets using this hashtag will be a daunting task. This is to show that microblog allow you to customize and categorise hashtags like categorizing them to UKZN.

Microblog Mining Vs Traditional Knowledge Management Mining

In microblog people can contribute their information and thoughts on one topic. Making the outcome of the discussion to be very fruitful. The discussion will allow the topic to be viewed in different angles by different people. If one is to mine this microblog discussion he will get rich information. Rather than the traditional knowledge mining the information is based on an individual’s effort. The information may be based on one angel or be looked in one dimension. Meaning it may be good but in one or two dimension. For an example when doctors are
commenting on the topic about mobile health in twitter. Each doctor may post his comment based on his practitioner background i.e. dentist, gynaecologist, physicians etc. Unlike when a dentist publishes his own research about mobile health, it will focus on dentists.

As microblogs offer service to post tweets in real time they have the power to keep recent and up-to-date information (Zhang and Sun, 2012). People tweet on each and every day. When mining a microblog one may enjoy this benefit. Unlike in traditional mining information may be outdated and irrelevant because it was gathered long ago. This is especially the case with information technology as it is evolving at a very high rate. When someone publishes an article in January you may find that it is no longer useful in November, because the situation might have been or have altered. This is the reason an antivirus always has to check latest updates for preventing malware because hackers are finding new ways of writing malware.

References are used in both methods, microblogs and in traditional methods. Those used in microblogs are URLs and the other uses traditional referencing. When mining data on a microblog and you need to view the reference you can easily click the URL and view it, just like that – it is simple. When using the traditional method you have to look at the reference page to know what article was referenced before you can spend some time searching the article online or in the library.

References


This document was written by

Lungisani Mgabisa

UKZN

210507284