Social Media and Corporate Data Warehouse Environments: New Approaches to Understanding Data

Debora S. Bartoo, Saint Joseph’s University, USA

ABSTRACT

This paper argues that organizations need to prepare for the integration of social media data into their data warehouses in order to fully understand their customers. Social media has quickly gained acceptance in its adoption and use and firms are eager to get their hands on it to better understand customer sentiment. However, social media data is different and more complex than traditional data and most data warehouses are not structured in a way for BI applications to easily make sense it. As a result, it is becoming critical for business intelligence teams to begin to understand the challenges this data presents and to better plan for the integration of this information into corporate data warehouses.

Keywords: Data Warehouse, Growth Trends, LSA, Privacy, Semantic Analysis, Sentiment Tracking, Social Media, Web Metrics

INTRODUCTION

Organizations seek ways to know more about their customers including behaviors, preferences, insights on new product launches and information related to products or services purchased. They want to understand what is being said on Internet blogs and gain a sense of customer sentiment. It is important to know if customer sentiment is positive or negative. This type of information can be gathered from small focus groups, but also can be captured from social media logs where customers are free to express their thoughts and views to anyone who will listen. The amount of data that can be analyzed from the Internet can’t compare to the limited data that can be gathered from focus groups.

The challenge is that the amount of data available from social media is incredible and continues to increase every day. “We create as much information in two days now as we did from the dawn of man through 2003” (Cukier, 2010).

Social media has been defined by Anthony Mayfield as “the collaborative, participatory technologies that allow people to communicate and collaborate” (Mayfield, 2010). Social media data is different from traditional data in data warehouses. Traditional data warehouses were built on a foundation of fixed length data fields that facilitate the extracting of information for analysis. Mining social media data becomes a challenge because of its volume and because it

DOI: 10.4018/jbir.2012040101
is in an unstructured format - text. Dealing with social media data generates the following questions: What if the data are not a fixed length? How do you measure conversational tone? How do you extract what the most common themes are that people choose to talk about? What key metrics should a dashboard contain to monitor activity? Is this even important to have business intelligence insights for social media? Some of these issues and insights need to be addressed now in order to be able to exploit the emerging social intelligence environment. A shift will need to occur in the traditional data warehouse approach to data to get ready to support the next generation of intelligence. This requires a different skill set from traditional structured data. Leading customer intelligence professionals are moving beyond listening to connecting social data with other forms of customer data and using this combined intelligence in both social and business interactions (Frankland, Khatibloo, & Stanhope, 2010).

“Social media is that which allows anybody to communicate with everybody.” In other words, it is consumer-generated content distributed through easy-to-access online tools (Sterne, 2010). Examples include You Tube, Facebook, LinkedIn, blogs, Twitter and many more. The web has become a focus group of experts who believe their views are important and want to be heard. Gaining insights into this data isn’t like methods used for traditional structured data. Social media data mining is more complex because it is about interactions. It is about conversations. It can happen with anyone anywhere across the world and, it happens instantly. It is unstructured. It doesn’t fit neatly into a data field. It doesn’t have pre-defined answers.

It is important to review foundational knowledge in web analytics to understand typical data that are already available and easily accessible. Web analytics require “tags” or pieces of JavaScript code to be embedded on each page. These “tags” provide an opportunity to gain insights into customer behaviors. For instance, with tags on a main page, analytics can be developed that show what the customer clicked on most frequently. A good example is in banking. The main web page contains these bits of code embedded and not visible to the visitor. If an online banking site has their login on the main page, it is likely the most clicked area on the page. Pages that are “tagged” can provide specific data related to that page including what elements a visitor clicked on and how often. This type of analysis can help organizations improve their site. It can also be used to gain a

The following section will examine what social media is and how it differs from traditional web analytics. It will provide insights into the challenges in comparison with traditional data warehouses. This will be followed by an analysis section that scrutinizes the implications and challenges to working with this new type of data.
Information Quality Assessment for Asset Management Systems
Sang Hyun Lee and Abrar Haider (2014). *Information Quality and Governance for Business Intelligence* (pp. 128-147).
[www.igi-global.com/chapter/information-quality-assessment-for-asset-management-systems/96148?camid=4v1a](www.igi-global.com/chapter/information-quality-assessment-for-asset-management-systems/96148?camid=4v1a)

A Study on a Combined Model in Business Intelligence for Improving Electronic Insurance
[www.igi-global.com/article/a-study-on-a-combined-model-in-business-intelligence-for-improving-electronic-insurance/132823?camid=4v1a](www.igi-global.com/article/a-study-on-a-combined-model-in-business-intelligence-for-improving-electronic-insurance/132823?camid=4v1a)