Summary

With the meteoric rise in popularity of social media, it has become a fundamental component in the success of a company. Though the use of social media brings with it a tremendous amount of unstructured and dynamic data. The article “Data warehouse design approaches from social media” discusses the approaches associated with dealing with this data and goes over basic data warehousing and social media concepts needed for these methods. The reference papers cited by the article fall under one of the following four categories. The first category is data warehouse concepts. These reference papers are used to explain basic data warehousing concepts such as fact tables and dimensions with the intent of making sure readers not as familiar with these concepts can still follow the contents of the article. Similarly the second category a reference article can fall under is social media concepts. These reference papers explain types of social media and other basic terms, so readers with different experiences are all on the same page. The third category of reference papers are the ones about “behavior analysis approaches”. Behavior analysis approaches concentrate on analyzing a user’s social media activity in order to develop innovative ideas based on this data. Reference papers for this category discuss different approaches proposed by a variety of experts. The fourth type of category is integration of sentiment analysis. This is similar to behavior analysis where it outlines a variety of data warehouse design approaches used to help decision makers. These four categories provide a clear overview of what is discussed in this article. To summarize, this article discusses two types of data warehouse approaches, behavior analysis and integration of sentiment analysis, and each author proposes a specific approach within those two types.

Motivation - what motivated authors to publish this paper

As time goes on the importance and influence of social media only continues to grow, proving it to not be a fad but a true staple of our social structure. Which makes the ability to extract user data from these social media platforms for the purpose of analysis even more important for companies. This importance is what motivated the authors to publish this paper, more specifically they wanted to write about the particular data warehouse design approaches used for social media. There were 11 approaches outlined within just this article, there are even more beyond this particular article’s scope. With this variety, making an informed decision about which approach is best for your needs is even more important. They wanted to write about the contributions of these design approaches, as well as their drawbacks. By providing this content at this level of detail, these authors have given those designing data warehouses (relating to social media) valuable information that will drive their decisions. Because as time goes on a company’s decisions will be even more informed by the data collected through social media use, so it becomes even more important to analyze these specific data warehouse design approaches.

Why - why is the paper's problem hard to solve?

The main problems are that social media is still relatively new and growing, there are many different types of social media, and that the amount of data in social media is enormous. All social media is varying in complexity and structure so they cannot all be built the same way - something like Twitter vs a Forum will need to be designed in their own unique way and cannot use the same solution. One example method proposed by Liu et al. was with what they called a 'social cube' - a text cube architecture; but this solution only worked for one type of social media and since social media contains a ton of information, it would never be suitable to work with such mass amounts of data. The other various solutions proposed by Rehman et al., Costa et al., and all other prospective ideas ended up with the same problems - the huge volume of data could not be properly represented and the solution only works with one specific type of social media.

Contribution - what does the paper contribute to the data warehousing community?

This paper has presented the main issues of trying to encompass all social media under one type of data warehouse. Most solutions only used one major social media such as Twitter or Facebook. The article came to the conclusion and theoretical solution that future studies should follow these steps: try to find a multidimensional model that can handle the diverse and vast amount of types of social media, then try to find an approach for the multidimensional model that can properly analyze the various user's opinions throughout social media.

Novelty - what is so novel about the idea?

There are many approaches on how you can utilize social media in many different methodologies. In order to demonstrate the multiple different strategies, they gave many examples of how people decided to approach the problem. There were many good ideas but I thought Rehman's approach was really well done and I think it's the best way to go about this in regards to twitter.

Validation - how did the author(s) validate the novelty

The author cited numerous examples to show how we could approach the problem of how to utilize these platforms effectively. There were five approaches cited and they also discussed some revisions the people made to their original approaches. There were some really good approaches but I thought Rehman's approach of extracting tweets and storing them in a data warehouse was a really good idea.

Works Cited

1. Moalla, I., Nabli, A., Bouzguenda, L., & Hammami, M. (2017). Data warehouse design approaches from social media: Review and comparison. *Social Network Analysis and Mining, 7*(1). doi:10.1007/s13278-017-0423-8