Week 5, Session 2

**Sharda: 5.1 - 5.5, 5.6 - 5.10; Articles: 1) Maximizing insight from unstructured data, 2) Start from the question - a guide to unstructured text analytics, 3) Fire up your social media strategy with big data analytics, 4) Getting to know your digital-age customers: an airline industry case study**

1. How does text mining differ from data mining?
2. What is NLP and how does it relate to text mining?
3. Explain the purpose of each step in the text mining process.
4. Pick an organization. Pick a product or service that this organization offers or plans to offer. What are some “core phrases”, “categories”, and “sentiments” that would be relevant for conducting sentiment analysis on reviews of the product or service? (Note: do not repeat the article’s examples)
5. Discuss some of the reasons why sentiment analysis engines produce inaccurate results.
6. For either stage 1 or stage 2 of the methodology proposed by the authors, a) describe the stage and b) discuss how the actions within the stage can improve sentiment analysis.
7. Pick an organization. Pick a product or service that this organization offers. Discuss the types of structured data that can be combined with sentiment analysis to gain greater insights into how this product or service is perceived in the marketplace.
8. In addition to sentiments and intentions, identify other insights that could be gained from text analytics. What would be some challenges to extracting these insights?
9. What is the difference between precision and recall? Discuss scenarios where precision would be more important than recall, and vice versa.
10. Explain and differentiate the following areas of web mining: content, structure and usage.
11. What are some strategies to improve the ranking of web pages in search engine results?
12. Discuss a few web analytics metrics that would be relevant for a specific company. How can these metrics help the company?
13. Is social media analytics relevant for all types of organizations? Why or why not?
14. Research and discuss either an aggregator or a software tool that can help organizations harvest social media data.
15. Explain with examples two techniques for “isolating the important conversations”.
16. Visit Kaggle.com. Search for competitions (active and completed) that are based on social media and unstructured data (e.g., Facebook, Twitter, reviews, sentiments, text, image, video). Briefly describe the competition including the sponsor, objective, reward, and datasets.
17. Discuss your own experiences with social media advertising and engagement. Are they effective?
18. Discuss some proactive and reactive strategies for avoiding or mitigating "negative virality" on social media that could be damaging to a company's image and brand.