**Project Proposal: Monitoring User-Generated Product Image**

***Xun Zha, Ao Pan, Henan Wang, Peiwei Li, and Yanran Liu***

**Assigned task:** *Using text mining of Amazon reviews to explore user-defined product*  
*highlights and issues (build the brand image).*

**Why Important**

After we figured out the user-defined product highlights and issues, the company of this product can improve the quality and user experience based on our text mining result for next generation of this product. In fact, they can mainly focus on to perfect customers need. If the text mining result would give the good feedback about that product, the company may use this to do the marketing campaign for the future.

**Is There Any Related Work**

A lot of survey companies provide the excel BI and text mining analysis based on the effective surveys. We expect small-size merchant companies who do not have the internal text mining team would like to use these third-party survey companies.

# Research Question Formulation

The ubiquitous product reviews provide firms with rich information on users’ opinions, but also challenges firms on immediately sensing those opinions. This project aims to address this issue by using text mining techniques. Specifically, we ask three research questions and plan to answer each of them as follows.

* What are most influential attributes that shape users’ perceptions of products? We plan to use LDA algorithm of topic modeling technique to find most influential attributes.
* How can we label top features that depict the merits and issues of each product by analyzing user-generated reviews? We plan to use TF-IDF method to find top features associated with either positive (e.g., 5-star) or negative (e.g., 1- and 2-star) reviews.
* Do users’ opinions on product attributes converge over time? We plan to calculate entropy of words distribution per product per time period, and check whether the entropy decrease over time.

# Task Decomposition and Allocation

* Data analysis
  + Prepare and clean data
    - Tokenization
    - Stopword Removal
    - Stemming and lemmatizing
  + Aggregate the top attributes associated with groups of products
  + Mine key words for each product to find out the most influential features to users
  + Run LDA model
  + Check the convergence of users’ opinions
* Final report written
  + Introduction
  + Methods & Analysis
  + Results
  + Discussion & Conclusion
* Presentation