

CUSTOMER REVIEW ANALYTICS:

- 1) Identify and dump product and vendor pages from several websites like flipkart,amazon etc.
 - a) Understand tools like scrapy,selenium etc.
 - i) Scrapy : contains mechanisms for crawling and scraping
 - ii) Selenium : contains mechanisms to render javascript and ajax enabled web pages
<http://stackoverflow.com/questions/17975471/selenium-with-scrapy-for-dynamic-page>
 - b) Get the sitemap for the site to be scraped or create one manually . Sitemaps give the overall structure of the website .
Ex : http://www.flipkart.com/sitemap_o_other_2015-08-17_1.xml.gz
Contains the many links that act like holders for a huge list of products .
(or)
The topmost menu in flipkart contains many categories like electronics,men,women etc . Now extract the links given in these menu items and you will find similar holders to many product links as described above while using sitemap .
 - c) Dump the product and vendor information pages locally.
- 2) Extract required product information from product pages in the dump .
 - a) Name
 - b) Rating
 - c) Vendor information
 - d) Customer reviews
- 3) For every product generate (key,value) pairs that indicate the aspect and the sentiment of the user for that particular aspect of the product .
- 4) Generate a product Hierarchy and extend the aspect sentiments to higher levels of hierarchy. For example , if we have aspect sentiments for “Samsung Galaxy On5” then one can first generate the terms - 1) “Samsung” 2) “Samsung Galaxy” etc from this product name and then extend the aspect sentiments to these higher levels of product categorisation . So, if battery life has negative sentiment for “Samsung Galaxy On5” then it will be reflected to an overall negative sentiment for “Samsung Galaxy” and “Samsung” as well . In this way, we can give a better picture of performance of the mobile for both the manufacturer as well as the user .

The part on how we actually perform aspect based sentiment analysis will be provided in the solution document .