**Deliverable D list of topics and substance you found**

The given Dataset has 220378 lines, each delivering a single comment

Below are its classifications

Dataset: 220378 comments

Cleaned Data: 192180

Cleaned Data with banks ID (tweeter): 32246

Cleaned Data without the above: 159934

Only BANK A – 41339 comments

Only bankB - 39332

Only bankC-22065

Only Bank D-45555

Comments containing two banks only:

A and B -2568

A and C- 705

A and D- 1094

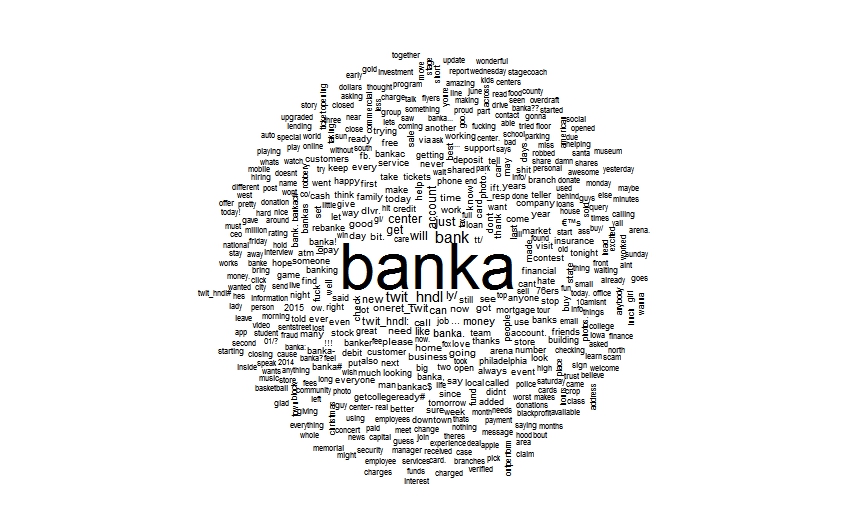
B and C- 64119

C and D -70036

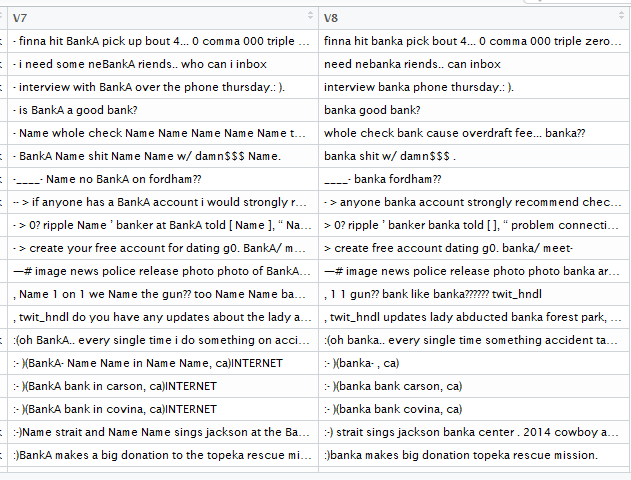
Comments containing all banks(ABCD)

1972 comments.

After processing Data pf Bank A(only)

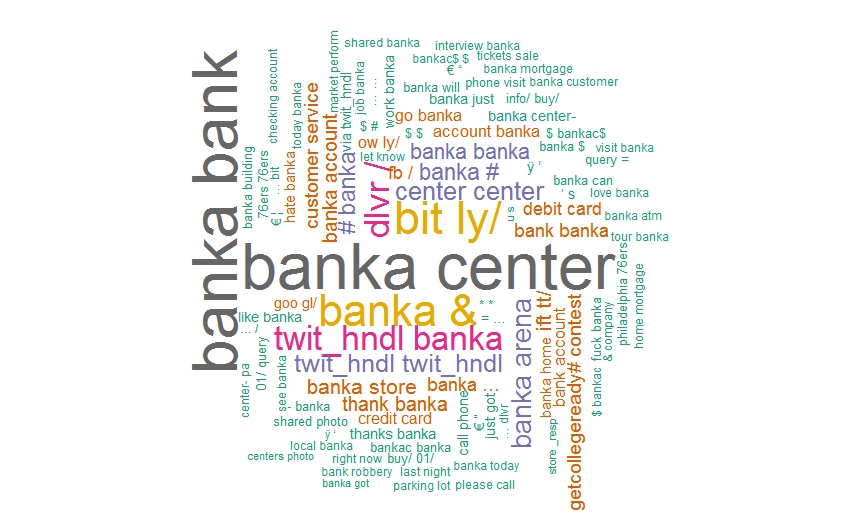
Below is the word cloud displaying high frequency words for bank ****

After Initial processing the bank A data and mining

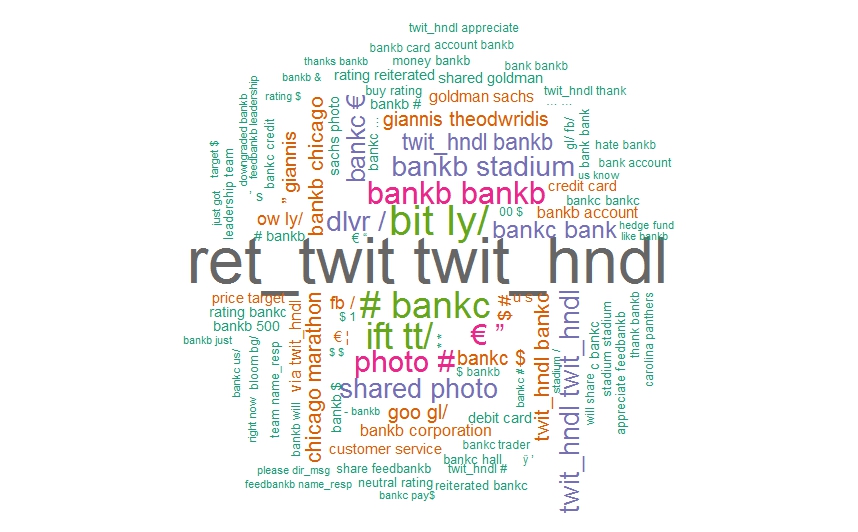
****

**Further, the symbols ☺ and ☺ are converted into happy and ☹ as sad and :@ as angry and the un necessary special characters are removed.**

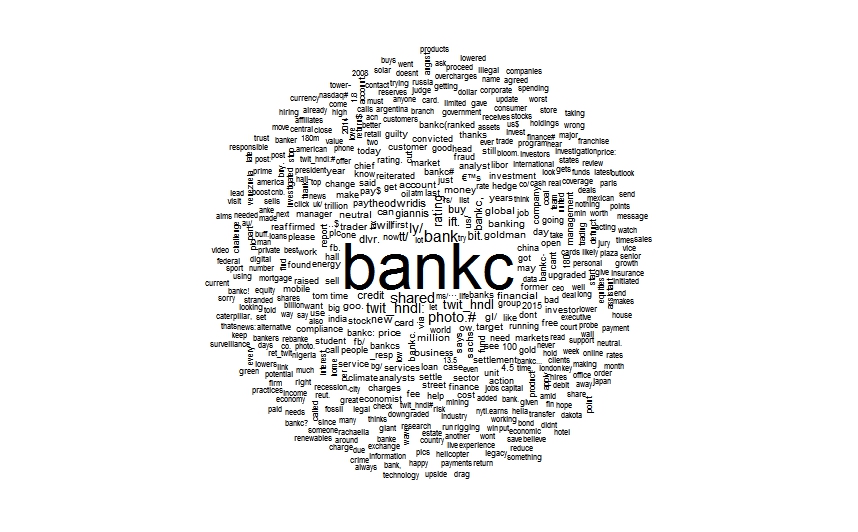
**Bigram world cloud gives us further more insight on the bank data.**

****

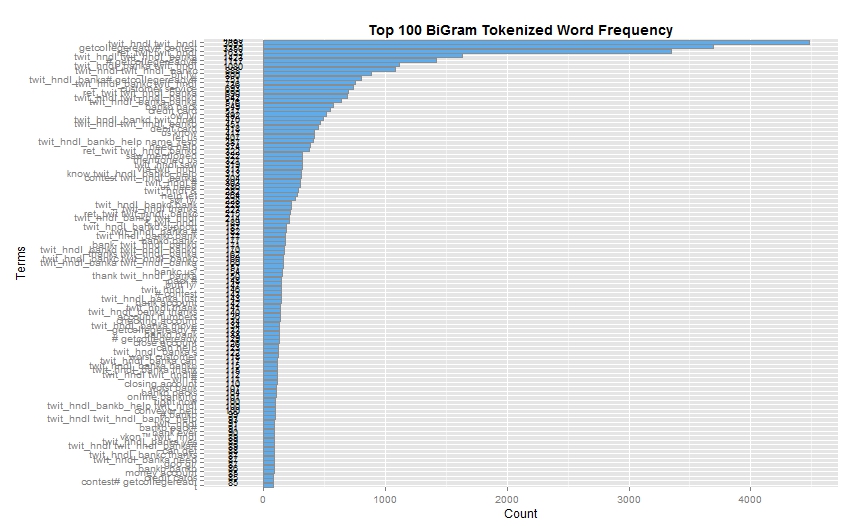
**From the data related to two banks**

****

**Similarly for other banks and tweeter data.**

****

**Tweeter Initial plot showing the tweet IDs in maximum comments**

****