

Ngram ReadMe

Introduction:

This part gives an array of mostly used phrases in review text of all businesses corresponding to the location they belong to.

Mapper: In Mapper class, I found all phrases that have two, three, four and five consecutive words in every review text. And then use a flag to decide if there is meaningless word (like I, you, they, in , on ,with etc.) in these phrases.

Output of Mapper class is <key: location number, value: phrase>

Reducer: In Reducer class, I want to find in every location, the mostly used phrases array ordered by their times of appearance from high to low.

I first store the output of Mapper class in a hash map where key is the phrase and value is the times the phrase appears in the given location. Every time the same phrase appears, I add one to the hash map value.

To order the phrases, I create an inner class “StringNode” with two fields: “word” and “count”. I add all keys and values in the hash map to an StringNode array. So that the keys of hashmap are stored in StringNode.word field and values of hashmap are stored in StringNode.count field.

Then I wrote a comparator for to sort the array by comparing the StringNode.count fields of all StringNode objects.

After sorting the array, the output of Reducer is <key: location number, value: array of top 20 mostly use phrases ordered by their times of appearance from high to low>

Input: ngramInput_clean.tsv (This file is from the line 37-38 in HiveRelatedCode in HiveRelated folder)

Output: ngramResult.tsv (The following is the result screenshot)

```
1 100-water park;customer service;cotton candy;rustlers rooste;go back;very nice;first
time;front desk;very good;check out;pretty good;arizona grand;prime rib;come back;lazy
river;not sure;fry bread;next time;frys electronics;great place;
2 105-root beer;glass bottles;glass bottle;flavored milks;heavy cream;egg nog;arizona
orange;fresh milk;just like;chocolate milk;comes from;local dairy;danzeisen dairy;farmers
market;like milk;so good;real milk;milk taste;flavors like;dairy itself;
3 113-carne asada;drive thru;asada fries;mexican food;carne asada fries;breakfast
burrito;pretty good;customer service;sour cream;breakfast burritos;rolled tacos;looks
like;super nachos;feel like;just didnt;5 stars;drive through;fast food;ive read;gone up;
4 114-mexican food;south phoenix;dr yocham;pollo fundido;customer service;very good;first
time;authentic mexican;ranch market;every time;can get;highly recommend;so good;take out;go
back;authentic mexican food;feel like;so much;years ago;know what;
5 115-grilled chicken;whole chicken;rice beans;grilled over;pinto beans;beans
tortillas;awesome sauce;side salad;rice beans tortillas;full chicken;doing low carb so;also
pack;1490 yupon;carb so;authentic good people too;other sides;salsa all;into these;bright
yellow;tiny bright;
6 116-customer service;mid k;not even;beauty supply;your manager;away from;new china;please
cashier;supply stores;never really;use your;actually help;best customer;new library;cashier
no;could not;mexican pastries;south mountain;much else;best customer service;
7 117-home depot;customer service;plant stand;arizona environmental;environmental
progress;arizona environmental progress;christmas tree;sales pitch;water softener;soft
water;your water;every year;no one;know what;sales tactics;great selection;your child;brand
new;water from;cant wait;
8 118-morning glory;south mountain;glory cafe;morning glory cafe;picnic tables;great
place;french toast;farm kitchen;outdoor seating;first time;mothers day;feel like;pecan
trees;very good;walk around;go back;check out;come back;pecan grove;monte cristo;
9 119-egg roll;fried rice;egg drop;drop soup;china restaurant;egg drop soup;so much;good
food;awakening seed;house chicken;cant beat;much sauce;lemon chicken;lunch specials;chicken
wings;very good;death certificate;loved one;take home;these items;
10 120-arizona mills;customer service;food court;az mills;first time;go back;mills mall;outlet
mall;come back;so much;too much;can get;ended up;fried rice;every time;great place;years
ago;chinese food;know what;not sure;
```