wb.data <- read.csv("d:/data/wb\_big/wbnew.csv");

//

index <- 1;

result.uid <- c(0);

result.mid <- c(0);

data <- wb.data;

length <- length(data[,1])

for(i in 1:length){

print(i);

temp <- as.character(data[i,3]);

temp <- strsplit(temp,split = ",")[[1]];

temp.length <- length(temp);

for(j in 1:temp.length){

result.uid[index] <- data[i,2];

result.mid[index] <- as.character(temp[j]);

index <- index + 1;

}

}

result <- data.frame("uid"=result.uid,"mid" = result.mid);

getMenction <- function(mention.size,data){

temp <- aggregate(result$uid,by = list("uid" = result$uid,"mid" = result$mid),length);

t <- temp[order(temp$x,decreasing = TRUE),];

colnames(t)[3] <- "weight";

t <- t[t$weight >= mention.size,];

}

wbStat <- function(max.size,data){

t <- data[data$weight >= max.size,];

node.v <- c(unique(t$mid),unique(t$uid));

node.v <- unique(node.v);

node <- data.frame("ID"=1:length(node.v),"UID"=node.v,"NAME"="","VALUE"=0,"CATEGORY"=3);

node$NAME <- as.character(node$NAME);

node[node$UID == "5242381821",5] = 1;

node[node$UID == "5242381821",3] = "奔跑吧兄弟官微";

node[node$UID == "5187664653",5] = 2;

node[node$UID == "5187664653",3] = "邓超";

node[node$UID == "1259193624",5] = 2;

node[node$UID == "1259193624",3] = "李晨";

node[node$UID == "1642351362",5] = 2;

node[node$UID == "1642351362",3] = "Baby";

node[node$UID == "1574684061",5] = 2;

node[node$UID == "1574684061",3] = "陈赫";

node[node$UID == "1275280670",5] = 2;

node[node$UID == "1275280670",3] = "郑凯";

node[node$UID == "1730330447",5] = 2;

node[node$UID == "1730330447",3] = "王祖蓝";

node[node$UID == "1426725707",5] = 2;

node[node$UID == "1426725707",3] = "包贝尔";

node[node$UID == "1254123322",5] = 2;

node[node$UID == "1254123322",3] = "王宝强";

write.csv(node,"d:/wb/node.csv",fileEncoding="UTF-8");

link <- merge(t,node,by.x = "mid",by.y = "UID",incomparables = NA);

link <- link[,1:4];

colnames(link)[4] <- "target"

link <- merge(link,node,by.x = "uid",by.y = "UID",incomparables = NA);

link <- link[,1:5];

colnames(link)[5] <- "source"

write.csv(link,"d:/wb/link.csv",fileEncoding="UTF-8");

category <- data.frame("INDEX"=c(1,2,3),"NAME"=c("跑男官微","跑男成员","其他人"),"KEYWORD"=c("跑男官微","跑男成员","其他人"),"BASE"=c("","",""))

write.csv(category,"d:/wb/category.csv",fileEncoding="UTF-8");

}

//微博树形结构处理

man <- c("5242381821","5187664653","1259193624","1642351362","1574684061","1275280670","1730330447","1426725707","1254123322");

tree.data <- data.frame("parent.uid"=c("","5242381821","5242381821","5242381821","5242381821","5242381821","5242381821","5242381821","5242381821"),"uid"=man,"parent.id"="","id"="","weight"=0);

t <- temp[(temp$uid %in% man),];

t <- t[(t$mid %in% man),];

tree.data$uid <- man;

for(i in 2:9){

tree.data[tree.data$uid==man[i],]$weight = t[t$uid == "5242381821" & t$mid==man[i],]$weight + t[t$mid == "5242381821" & t$uid==man[i],]$weight;

}

man.2 <-man[2:9];

t.2 <- temp[(temp$uid %in% man.2),];

t.2.mid <- temp[(temp$mid %in% man.2),];

t.2 <- rbind(t.2,t.2.mid)

t.2 <- unique(t.2)

t.2 <- t.2[t.2$uid != "5242381821",]

t.2 <- t.2[t.2$mid != "5242381821",]

t.2 <- t.2[t.2$mid != t.2$uid,]

data3 <- c(unique(t.2$mid),unique(t.2$uid))

data3 <- unique(data3)

data3 <- data3[-(data3 %in% man.2)]