$ cat enwiki-latest-stub-articles8.xml | ./mapper.R | sort | ./r

educer.R > output\_wReducer.txt

\*\*\* caught segfault \*\*\*

address 0x7f22d6d4a460, cause 'memory not mapped'

Traceback:

1: ls(env, all = TRUE)

aborting ...

Segmentation fault (core dumped)

**FIX**

if the data comes in sorted we don’t need to keep a list of values.

word\_count <- 0

prev\_word <- ""

con <- file("stdin", open = "r")

while (length(line <- readLines(con, n = 100, warn = FALSE)) > 0) {

val<-strsplit(line, "\n")

for(i in seq(val)){

line <- trimWhiteSpace(val[[i]])

split <- splitLine(line)

word <- split$word

count <- split$count

if(word==prev\_word){

word\_count<-word\_count+count

} else {

if(prev\_word!="") cat(prev\_word,"\t",word\_count,"\n",sep="")

prev\_word<-word;

word\_count<-count;

}

}

if(word!="") cat(word,"\t",word\_count,"\n",sep="")

}

close(con)