**Markov Chain explanation**

I think the most important part of Markov Chains, for our task, is that history doesn't matter. All that matters is the current state. The simplest possible model is that the current state is simply the current N-Grams, for some chosen N. A more complicated model would take into account sentence position or part of speech tagging to define the current state. Basically, your prediction depends on the current state, and nothing else.

So you have to come up with a way of defining your transition possibilities. That is, given the current state, what are the odds you'll go to the possible set of states that are legal from the current state. Again, in the simplest possible model, the current state is the most recent NGram (pick an N, 1, 2, 3... whatever) and the "transitions" are the next word choices. In a slightly more complicated model, the current state is the current 1-gram, 2-gram, 3-gram, and you use a backoff model to measure the transitions possibilities, as suggested on the [Task 3](https://class.coursera.org/dsscapstone-004/wiki/Task_3) page .

In my opinion, we don't care to measure every single transition probability. Because we only care to predict the **most likely** next word or words. I hope this helps.

**Dealing With Garbage characters – Unicode to ANSI?**

Thank you guys for pointing me to the right function.  
If someone else have this problem you can check your encoding options on your system with iconvlist().  
After a lot of trials the only thing that worked for me at my system was -   
t<-iconv(t, to="ASCII", sub = "")  
done right after reading the data with readLines() .