

# Text Mining the Hitchhiker Trilogy

*Greg Johnson*

Let's read in Doug's first textbook.

```
#setwd("~/Documents/Analytics/R/DouglasAdams")
book1 <- readLines("data/HitchHiker.txt")
head(book1,20)
```

```
## [1] ""
## [2] ""
## [3] "
## [4] "
## [5] "
## [6] "
## [7] "
## [8] "
## [9] "
## [10] ""
## [11] ""
## [12] "
## [13] "
## [14] "
## [15] ""
## [16] "
## [17] "western spiral arm of the Galaxy lies a small unregarded yellow sun."
## [18] "
## [19] "utterly insignificant little blue green planet whose apedescended life"
## [20] "forms are so amazingly primitive that they still think digital watches are"
```

Our data are in! But they're not tidy in the sense that we want one token per line. For a first look at our data, we will look at words.

```
3
```

```
## [1] 3
```