

**Steps**

**Data**

a. Break title into list of chars, regrouping into list of words, attaching reference numbers.

```
"It is easier to fight for one's principles ..."
"The only normal people are the ones you ..."
"Love thy neighbor as thyself, but choose ..."
```

```
(("It" "is" "easier" ...) 0)
(("The" "only" "normal" ...) 1)
(("Love" "thy" "neighbor" ...) 2)
```

b. Split numbered titles at word boundaries.

```
(() ("It" "is" "easier" ...) 0)
(("It") ("is" "easier" ...) 0)
(("It" "is") ("easier" ...) 0)
. . .
(() ("The" "only" "normal" ...) 1)
(("The" ) ("only" "normal" ...) 1)
(("The" "only")("normal" ...) 1)
. . .
(() ("Love" "thy" "neighbor" ...) 2)
(("Love") ("thy" "neighbor" ...) 2)
. . .
```

c. Drop entries keyed by noise words.

```
((("It" "is") ("easier" ...) 0)
. . .
(("The" ) ("only" "normal" ...) 1)
(("The" "only")("normal" ...) 1)
. . .
(() ("Love" "thy" "neighbor" ...) 2)
(("Love") ("thy" "neighbor" ...) 2)
. . .
```

d. Sort split titles according to second halves.

```
((("It" "is") ("easier" ...) 0)
(() ("Love" "thy" "neighbor" ...) 2)
(("The" "only")("normal" ...) 1)
(("The" ) ("only" "normal" ...) 1)
(("Love") ("thy" "neighbor" ...) 2)
. . .
```