

# Re: handling text

---

*Source:* <http://coding.derkeiler.com/Archive/Fortran/comp.lang.fortran/2006-01/msg00860.html>

---

- *From:* glen herrmannsfeldt <[gah@xxxxxxxxxxxxxxxxxxx](mailto:gah@xxxxxxxxxxxxxxxxxxx)>
  - *Date:* Sat, 28 Jan 2006 18:51:03 -0800
- 

Dr Ivan D. Reid wrote:

(snip)

I'd prefer to use awk or one of its successors such as perl; with awk you can take each word of an input string and increment a count for that word, creating the array element with a default initial value of 0 if that word hasn't been encountered before. Something like:

```
{ for (i=1; i<=NF; i++) {count[$i]++;}
}
END {for (word in count) {print word,count[word];}
}
```

Yes, I like awk for this type of problem.

The important part is a hash table, which I don't believe that Fortran includes yet.

Last time I did work counting I started in awk, but when that was too slow (I had 15GB to count) I went to Java. Using StringTokenizer to find words and Hashtable keep track of them it isn't so hard. With JIT on it runs reasonably fast. (It took some hours to do 15GB, though computers are faster now.)

If you have a hashtable routine for Fortran it probably isn't so hard to do.

-- glen

Re: handling text