

---

Subject: Re: comparing and concatenating arrays...please help!!

Posted by [Michael Wallace](#) on Thu, 08 Jan 2004 21:22:40 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

> Hi Pepijn,

>

> Many thanks for your reply.

>

> My files are \_huge\_. File 1 is about 250000 lines and file2 about half

> that. which is why my code was taking about 4 days!

>

> I'll give the sorting a try, but if anyone else has any suggestions,

> they'll be gratefully received. I'm not a newbie, but am daunted by

> some of the operations you suggested below... I certainly take

> inspiration from all you guys though!

I'd try the sorting attack myself -- sort first and then merge the sorted data. I don't know what kind of memory and processing constraints you're under, so this might not be applicable if you have sufficient memory. If you run into a memory problem trying to sort this much data, you can divide the sort into chunks. Read and sort the first X many lines, then read and sort the next X many, and so on. Then take all the chunks and merge them. Finally, merge of the sorted file1 and sorted file2. While there will be more operations than the first approach, the running time will still be on the order of  $n \log(n)$  which is still better than your original running time of  $n^2$ , and you will chew up a lot less memory. Again, this is just a suggestion if you run into memory problems trying to do such a large sort.

Mike W

---