

Tip 30: Use Apophenia to read in data and configuration info

Ben Klemens

30 November 2011

level: Basic for anybody dealing with data

purpose: Use libraries!

This is a special case of Tip #2 (Entry #051) about using pre-existing libraries wherever possible. After all, C's big edge is that it's been around for forty years; that's a lot of time for useful libraries to get written.

Reading in text is an especially difficult problem that everybdoy has to deal with so it is especially library-appropriate. Despite my self-conscious desire to not self-promote, I'm gonna tell you that Apophenia does a decent job with this.

First, let's generate a data set. I'll wrap it up as a here document, as per Tip # 8 (Entry #058), so you can just paste this onto the command line:

```
cat > text_data << "."
left|middle|right
2|5| 12
3|8|9
3|8|Galia est omnis divisa en partes tres
.
```

The sample data shows the first tip for the day: use pipes as field delimiters. Pipes really look like the bounds between fields, and they rarely appear in the data you're putting into a text file. The default for so many systems is commas or tabs, both of which are just asking for glitches.

Reading a data set to a matrix is pretty trivial via Apophenia. In this example, I'll stretch it out by first reading into the database (instead of directly using `apop_text_to_data`, which would save two lines of code but lose the non-numeric input). [And remember Tip #9 (Entry #059) about compiling C code via here document? It's how I test all the sample code I put here, and is still an easy way for you to try it all out.]

```
#include <apop.h>

int main() {
    sprintf(apop_opts.input_delimiters, "|");
    apop_text_to_db("text_data", "datatab");
}
```

```
    apop_data *indata = apop_query_to_data("select "
                                           "left, middle from datatab");
    Apop_col(indata, 0, firstrow);
    Apop_col(indata, 1, secondrow);
    printf("first column sum: %Lg\n", apop_sum(firstrow));
    printf("second column sum: %Lg\n", apop_sum(secondrow));
}
```

If you installed the Apopenia library, then you also have the command-line `apop_text_to_db`, which just runs the C function in the second line of `main`.