

Re: text parsing

Source: <http://coding.derkeiler.com/Archive/PHP/comp.lang.php/2008-01/msg01889.html>

- *From:* The Natural Philosopher <a@xxx>
 - *Date:* Wed, 23 Jan 2008 15:30:54 +0000
-

Carolyn Marenger wrote:

McKirahan wrote:

"Carolyn Marenger" <cajunk@xxxxxxxxxxxx> wrote in message
[news:7c0f\\$4795ea54\\$cf70133e\\$1079@xxxxxxxxxxxx](mailto:news:7c0f$4795ea54$cf70133e$1079@xxxxxxxxxxxx)

McKirahan wrote:

"Carolyn Marenger"
<cajunk@xxxxxxxxxxxx> wrote in message
[news:74fb1\\$479501d1\\$cf70133e\\$7458@xxxxxxxxxxxx](mailto:news:74fb1$479501d1$cf70133e$7458@xxxxxxxxxxxx)

Can someone point me in
the direction of some good
documentation on

text

parsing?

I want to take a bunch of
text files (rtf), read them in
and dump the
contents in a database. The
files are effectively a flat
file

database,

with I suspect some fairly
intricate programming
needed to process the
files. Unfortunately, they are
laid out for human
readability, not

data

Re: text parsing

conversion.

A few questions.

How many is a "bunch"?

What would the target database be —
MySQL?

What table and column structures do you
envision?

Perhaps simply a single table with two
columns:

filename (key) and a memo field containing
the data?

What is the purpose behind doing this?

A few answers

A bunch is about a dozen. Basically one large file that was
broken into
sixteen subsets, following the initial letter for each record.

The target database would be MySQL

I haven't looked too closely at the data, but I think one main
table
with a few linked tables for those cases where there may be
more than
one piece of data for a category. There are about 25
categories to each
record. Eventually there would be additional structure added
around the
imported data, but that isn't relevant to importing the data
itself. (I
will confirm this before beginning to code.

The purpose: I am a D&D fan and I run games. I would like
to be able to
reference the material and automate much of the process so I
don't have
to lug and reference 20lbs of books.

Any chance the RTF files are online so I could look at them?

Perhaps <http://www.wizards.com/default.asp?x=d20/article/srd35?>
<http://www.wizards.com/d20/files/v35/SRD.zip> contains 88 RTF files.

Also, I gather, this might be a one-time effort; correct?

Re: text parsing

Not what you requested but ...

I've developed a VBScript solution that takes the following approach: for a given folder, each RTF file is opened in MS-Word and saved as a text file which is opened and read then saved in an MS-Access database table containing 3 columns: id (AutoNumber), file, data.

Using those 86 RTF files it created a 10MB MS-Access database.

Yes, they are online. Yes, you can look at them. Yes, those are the files except I only care about the 16 monster files. Yes, this is a one time effort.

My goal is to create an encounter generation program – where I key in climate, geography, season, encounter level, time of day, proximity to civilization, and the application gives me a suggested random encounter suited to the scenario. For example, if the party was wandering around the city sewers on a hot summer night, they might encounter a pack of giant rats being led by a were rat.

Only if

1/. It was los angeles

2/. They had all taken too many mind enhancing drugs.

Otherwise its likely to be Viles disease, at the most interesting ;-)

I would then want the program to determine how many rats, how many hit points each, and any other pertinent variable data, including what weapons and treasure the wererat was carrying and using.

Having the rtf's loaded into a database like your script does, would enable faster searches, it would not go the next step and perform the various calculations based on the results of the searches. It is a good start, but if it has stripped any of the rtf encoding, it may make it harder to have a script find the various 'fields'.

Go full database surely. The art is to define the 'monster' table with extensibility for all the monster classes one might encounter.

When doing ANYTHING based on a database, the most important thing is to spend time designing table layouts. And write a data dictionary. And keep it up to date.

Re: text parsing

Re: text parsing

Thanks, Carolyn