Subject: Retriving data in record containing ""
Posted by forlano on Sat, 03 Sep 2011 14:47:11 GMT

Hello,

I have records like this

","201113","2577","173","2593","GM","4600010","GER "

I need to retrive the data separated by ','. Unfortunately this sign is used inside column 4 and give problem to Split(). Column 4 sometime has one ',' and sometime two ','.
Is there some special method to process such a record?

Thanks,
Luigi

Subject: Re: Retriving data in record containing ""
Posted by tojocky on Sat, 03 Sep 2011 17:16:24 GMT

Hi,

It is very simple:
0. if first char is new line then create new record
1. first check if the next char is "
2. if first char is " yes then get value till next ", skip "," and repeat 1
3. if first isn't " then get next value till ,, skip "," and repeat 1

in your example, point 3 is never used.

Hope this help you.
Ion.

forlano wrote on Sat, 03 September 2011 17:47Hello,

I have records like this

","201113","2577","173","2593","GM","4600010","GER "

I need to retrive the data separated by ','. Unfortunately this sign is used inside column 4 and give problem to Split(). Column 4 sometime has one ',' and sometime two ','.
Is there some special method to process such a record?

Thanks,
Luigi

Subject: Re: Retriving data in record containing ""
Posted by forlano on Sat, 03 Sep 2011 17:52:25 GMT

tojocky wrote on Sat, 03 September 2011 19:16Hi,

It is very simple:
0. if first char is new line then create new record
1. first check if the next char is "
2. if first char is " yes then get value till next ", skip "," and repeat 1
3. if first isn't " then get next value till ,, skip "," and repeat 1

in your example, point 3 is never used.

Hope this help you.
Ion.

Hi Ion,

I hoped in something at level of Split().
What I done was to replace '",' with '|' and obtain a string suitable to be used with Split, then remove again the '"' in the substring.
Perhaps your solution is better.

Luigi

Subject: Re: Retriving data in record containing ""
Posted by sergeynikitin on Sat, 03 Sep 2011 19:29:51 GMT

This very slow. I recomend use Regular Expression (package plugins/pcre)

CSV-parser (regex pattern below)
^(("(?:[^"]|"")*"|[^,]*)(,("(?:[^"]|"")*"|[^,]*))*)$

or

And in other words:
http://www.google.com/search?client=ubuntu&channel=fs&q=retrieving+CVS+regular+expression&ie=utf-8&oe=ut f-8

Subject: Re: Retriving data in record containing ""
Posted by forlano on Sun, 04 Sep 2011 08:19:03 GMT
View Forum Message <> Reply to Message

sergeynikitin wrote on Sat, 03 September 2011 21:29This very slow. I recomend use Regular Expression (package plugins/pcre)

Prevet Sergey,

I agree it is slow. I tried your pattern
#include <Core/Core.h>
#include <plugin/pcre/pcre.h>

using namespace Upp;

```
CONSOLE_APP_MAIN
{
 String ss, s = LoadFile("in.txt");
 String p1 = "^((\"(?:[^\"]|\"\")*\"|[^,]*)(,(\"(?:[^\"]|\"\")*\"|[^,]*))*)$";
  RegExp r0(p1);
  r0.Match(s);
 Vector<String> field=r0.GetStrings();
  for (int i=0; i<field.GetCount(); i++) {
    ss << field[i] << "\n";
 }
 SaveFile("out.txt", ss);
}
```

and in.txt is
"Luigi", "Ivan, Petrovic", "43434"
but it give me weird output. I would like to get the data inside ".

Luigi
Ivan, Petrovic
43434

Perhaps I need to modify the pattern as I have 14 fields
http://www.kimgentes.com/worshiptech-web-tools-page/2008/10/
14/regex-pattern-for-parsing-csv-files-with-embedded-commas- dou.html

Luigi

Subject: Re: Retriving data in record containing ""
Posted by mirek on Mon, 05 Sep 2011 09:19:14 GMT

View Forum Message <> Reply to Message

forlano wrote on Sat, 03 September 2011 10:47Hello,

I have records like this

","201113","2577","173","2593","GM","4600010","GER "

I need to retrive the data separated by ','. Unfortunately this sign is used inside column 4 and give problem to Split(). Column 4 sometime has one ',' and sometime two ','.
Is there some special method to process such a record?

Thanks,
Luigi

If those strings are C-like, use CParser.

If it is Csv, I happen to have some routines (not in U++ sources) for that as well...

Mirek

Subject: Re: Retriving data in record containing ""
Posted by forlano on Mon, 05 Sep 2011 10:52:25 GMT

View Forum Message <> Reply to Message

mirek wrote on Mon, 05 September 2011 11:19forlano wrote on Sat, 03 September 2011 10:47Hello,

I have records like this

","201113","2577","173","2593","GM","4600010","GER "

I need to retrive the data separated by ','. Unfortunately this sign is used inside column 4 and give problem to Split(). Column 4 sometime has one ',' and sometime two ','.
Is there some special method to process such a record?

Thanks,
Luigi

If those strings are C-like, use CParser.

If it is Csv, I happen to have some routines (not in U++ sources) for that as well...

Mirek

Yes, it a CSV file (the same that gave problem with accented character).
Luigi

Subject: Re: Retriving data in record containing ""
Posted by mirek on Mon, 05 Sep 2011 12:03:58 GMT
View Forum Message <> Reply to Message

Currently I am using this to parse csv:

```
Vector<String> ReadCsvLine(Stream& s, int separator, byte charset)
{
 Vector<String> r;
 bool instring = false;
 String val;
 for(;;) {
  int c = s.Get();
  if(c == '\n' || c < 0) {
   r.Add(ToCharset(CHARSET_UTF8, val, charset));
   return r;
  }
  else
  if(c == separator && !instring) {
   r.Add(ToCharset(CHARSET_UTF8, val, charset));
   val.Clear();
  }
  else
  if(c == '\"') {
   if(instring && s.Term() == '\"') {
    s.Get();
    val.Cat('\"');
   }
   else
    instring = !instring;
  }
  else
  if(c != '\r')
   val.Cat(c);
```

```
 }
}
```