Subject: Re: Is there a token function?
Posted by mirek on Sat, 22 Apr 2006 20:39:01 GMT

Well, if you are not to process 100MB+ files that are not well divided into lines and has similar lexical syntax to C, the easiest and fastest U++ way is to LoadFile into String and then use CParser.

If you DO want to process 100MB+ files and/or they are divided into lines, usually you can use similar approach, just use CParser per each single line. Or process it in any other way - in this particular case, I believe "Split" would be the best option (as you have fields separated by ';').

BTW, what you present is kind of import I have to provide for my customers quite often. Skeleton of such parser usually looks like

```
FileIn in(path_to_file);
in.GetLine(); // to eat "header" line
try {
  while(!in.IsEof()) {
    Vector<String> field = Split(in.GetLine(), ';');
    if(field.GetCount() != 11)
      throw CParser::Error("");
    String name = field[0];
    String coutnry = field[1];
    CParser p(field[2]);
    Date d;
    d.day = p.ReadInt();
    p.PassChar('.');
    d.month = p.ReadInt();
    p.PassChar('.');
    d.year = p.ReadInt();
    if(d.year < 20)
      d.year += 2000;
    else
    if(d.year < 100)
      d.year += 1900;
  }
}
catch(CParser::Error) {
// invalid file
}
```