Reading, Writing, and Parsing Text Files Using C++
(Updated)

by Robert J Yager
NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer’s or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.
Reading, Writing, and Parsing Text Files Using C++
(Updated)

Robert J Yager
Weapons and Materials Research Directorate, ARL
This report presents a set of functions, written in C++, that can be used to read, write, and parse text files. Updates include bug fixes and the addition of a new function.
Contents

Acknowledgments iv

1. Introduction 1

2. Background 1

3. Reading and Writing Text Files 1
   3.1 ReadTextFile() Function .................................................................1
   3.2 WriteTextFile() Function ...............................................................2
   3.3 ReadTextFile()/WriteTextFile() Example.........................................3

4. Comment Removal 3
   4.1 RemoveLineComments() Function .................................................3
   4.2 RemoveBlockComments() Function...............................................4

5. Parsing Text 5
   5.1 Parse() Function ...........................................................................5
   5.2 Parse2D() Function ........................................................................6
   5.3 PreParse() Function ......................................................................8

6. Summary 9

7. References 11

Distribution List 12
Acknowledgments

I would like to thank Dr Benjamin Breech of the US Army Research Laboratory’s Weapons and Materials Research Directorate. Dr Breech provided technical and editorial recommendations that improved the quality of this report.

I would also like to thank Captain Benjamin Flanders of the US Army. Captain Flanders pointed out errors in my original code and helped develop new code.
1. **Introduction**

This report presents a set of functions, written in C++, that can be used to read, write, and parse text files.

A summary sheet is provided at the end of this report. It presents the yIo2 namespace, which contains the ReadTextFile(), WriteTextFile(), RemoveLineComments(), RemoveBlockComments(), Parse(), Parse2D(), and PreParse() functions.

2. **Background**

I previously published a report titled *Reading, Writing, and Parsing Text Files Using C++*, which was very similar to this report. Unfortunately, the code presented in that report contained several errors. This report fixes those errors and introduces several other improvements. A new function, PreParse(), has also been added.

The following changes from the previous code affect function behavior:

- The default value for the WriteTextFile() function’s mode parameter has been changed from "w" to "wb".
- When using the default parameters, the RemoveBlockComments() function previously interpreted "*/" as a complete block comment.
- The size of the output vector of the Parse() function was previously never less than 1, even if the function was passed a pointer that pointed to a NULL character.
- The Parse2D() function previously could only parse text that ended with a delimiter.

3. **Reading and Writing Text Files**

3.1 **ReadTextFile() Function**

The ReadTextFile() function can be used to create a character array that contains all of the information from a text file.

Note that the ReadTextFile() function uses the new[] operator to allocate memory for the character array that is pointed to by its return value. Thus, to avoid memory leaks, each use of the ReadTextFile() function should be accompanied by a use of the delete[] operator.
ReadTextFile() Code

```c
inline char*ReadTextFile(const char*f){
    FILE*F=fopen(f,"rb");
    if(!F)printf("Can't find \"%s\".\n",f),exit(1);
    unsigned n=fseek(F,0,SEEK_END),n=ftell(F),rewind(F);
    char*s=new char[n+1];
    fread(s,1,n,F),fclose(F),s[n]=0;
    return s;
}
```

ReadTextFile() Parameters

- **f** points to the name of a text file.

ReadTextFile() Return Value

The ReadTextFile() function returns a pointer to the beginning of a character array that stores all of the information from the file specified by the input parameter `f`.

If the file that is specified by `f` cannot be opened, the ReadTextFile() function calls the `exit()` function with status code 1. Inability to open a file is typically the result of an incorrectly specified filename or path.

### 3.2 WriteTextFile() Function

The WriteTextFile() function can be used to write text to a file.

WriteTextFile() Code

```c
inline unsigned WriteTextFile(const char*f,const char*s,const char*m="wb"){ 
    FILE*F=fopen(f,m);
    if(!F)printf("Can't open \"%s\".\n",f),exit(1);
    unsigned n=fwrite(s,1,strlen(s),F);
    fclose(F);
    return n;
}
```

WriteTextFile() Parameters

- **f** points to the name of the text file that will be written to.
- **s** points to the text that will be written to the text file.
- **m** specifies how text will be written to the output file. Use "wb" to overwrite an existing file. Use "ab" to append to the end of an existing file. In either case, if a file with the same name as the output filename doesn’t already exist, a new file will be created. The default value is "wb".
WriteTextFile() Return Value

The WriteTextFile() function returns the number of characters that were successfully written to the output file.

If the file that is specified by f cannot be opened, the WriteTextFile() function calls the exit() function with status code 1. Inability to open a file is often the result of an incorrectly specified filename or path. However, it can also be the result of a file being marked as read only, as may be the case if a file is open in another program.

3.3 ReadTextFile()/WriteTextFile() Example

The following example uses the WriteTextFile() function to create a file named "example.txt". The ReadTextFile() function is then used to read the newly created file.

```cpp
#include "y_io_2.h" //......................................yIo2,<cstdio>{printf}
int main() { //<=====SIMPLE EXAMPLE FOR WriteTextFile() & ReadTextFile() FUNCTIONS
    yIo2::WriteTextFile("example.txt","Hello World!\n");
    char*s=yIo2::ReadTextFile("example.txt");
    printf("\%s",s);//........................display the contents of "example.txt"
    delete[]s;
} //~~~~~~YAGENAUT@GMAIL.COM~~~~~~~~~~~~~~~~~~~~~~~~~LAST~UPDATED~20JUL2014~~~~~~
```

OUTPUT:

```
Hello World!
```

4. Comment Removal

4.1 RemoveLineComments() Function

The RemoveLineComments() function can be used to overwrite line comments. A line comment is a comment that begins with some identifying set of characters and continues to the end of the line. Memory that is occupied by line comments is not actually freed by the RemoveLineComments() function. Instead, each line-comment character is replaced with a user-specified character.

RemoveLineComments() Code

```cpp
inline char*RemoveLineComments(//<=====================OVERWRITE LINE COMMENTS
    char*s,//<---------------------------TEXT THAT MIGHT CONTAIN LINE COMMENTS
    const char*c="#",//<------------------------BEGINNING-OF-COMMENT INDICATOR
    char r=' '){//<------THE CHARACTER THAT WILL BE USED TO OVERWRITE COMMENTS
    for(char*t=s;t=strstr(t,c);memset(t,r,strcspn(t,"\n\f\r")));
    return s;
} //~~~~~~YAGENAUT@GMAIL.COM~~~~~~~~~~~~~~~~~~~~~~~~~LAST~UPDATED~20JUL2014~~~~~~
```
RemoveLineComments() Parameters

s  s points to text that will have its line comments overwritten.

c  c points to text that specifies the beginning of a line comment. The default value is "#".

r  r specifies the character that will be used to replace the characters in the line comment. The default value is a space.

RemoveLineComments() Return Value

The RemoveLineComments() function returns the input pointer s.

4.2 RemoveBlockComments() Function

The RemoveBlockComments() function can be used to overwrite block comments. A block comment is a comment that begins with some identifying set of characters and ends with a (potentially) different set of identifying characters. Block comments may or may not span multiple lines. Memory that is occupied by block comments is not actually freed by the RemoveBlockComments() function. Instead, each block-comment character is replaced with a user-specified character.

If the RemoveBlockComments() function encounters a beginning-of-comment indicator, but not an end-of-comment indicator, then the entire comment is ignored. The RemoveBlockComments() function does not recognize nested block comments as being nested.

RemoveBlockComments() Code

```c
inline char*RemoveBlockComments(  //<==============OVERWRITE BLOCK COMMENTS
    char*s,//<--------------------------TEXT THAT MIGHT CONTAIN BLOCK COMMENTS
    const char*c="/*",//<-----------------------BEGINNING-OF-COMMENT INDICATOR
    const char*e="*/",//<-----------------------------END-OF-COMMENT INDICATOR
    char r=' '){  //<------THE CHARACTER THAT WILL BE USED TO OVERWRITE COMMENTS
    int m=strlen(c),n=strlen(e);
    for(char*t=s,*u;t=strstr(t,c);memset(t,r,u-t+n))if(!(u=strstr(t+m,e)))break;
    return s;  //...........block comments without end indicators aren't overwritten
}  //~~~YAGENAUT@GMAIL.COM~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~LAST~UPDATED~20JUL2014~~~~~~
```

RemoveBlockComments() Parameters

s  s points to text that will have its block comments overwritten.

c  c points to text that indicates the beginning of a block comment. The default value is "/\*".

e  e points to text that indicates the end of a block comment. The default value is "*/".
**r** specifies the character that will be used to replace the characters in the line comment. The default value is a space.

### RemoveBlockComments() Return Value

The RemoveBlockComments() function returns the input pointer \( s \).

### RemoveLineComments()/RemoveBlockComments() Example

The following example code uses the RemoveLineComments() and RemoveBlockComments() functions to overwrite comments in a sample block of text.

```cpp
#include "y_io_2.h"  //......................................yIo2,<cstdio>{printf}
int main(){  //<===SIMPLE EXAMPLE FOR RemoveLineComments() & RemoveBlockComments()
  char s[]="#sample line comment\nHello/*sample block comment*/World!\n";
  printf("ORIGINAL TEXT:\n%s\n",s);
  printf("LINE COMMENTS REMOVED:\n%s\n",yIo2::RemoveLineComments(s));
  printf("LINE & BLOCK COMMENTS REMOVED:\n%s\n",yIo2::RemoveBlockComments(s));
}  //~~~~~~YAGENAUT@GMAIL.COM~~~~~~~~~~~~~~~~~~~~~~~~~LAST~UPDATED~20JUL2014~~~~~~
```

**OUTPUT:**

<table>
<thead>
<tr>
<th>ORIGINAL TEXT:</th>
<th>LINE COMMENTS REMOVED:</th>
<th>LINE COMMENTS AND BLOCK COMMENTS REMOVED:</th>
</tr>
</thead>
</table>
| #sample line comment
Hello/*sample block comment*/World! | Hello/*sample block comment*/World! | Hello World! |

---

### Parsing Text

#### 5.1 Parse() Function

The Parse() function can be used to separate the text into tokens. Tokens are blocks of text that are separated by delimiting characters. Common delimiters are commas, tabs, and spaces.

The Parse() function works by searching for tokens. When a token is found, the delimiting character that immediately follows the token is replaced by the NULL character, and a pointer to the beginning of the token is stored in a vector. Any set of consecutive delimiter characters are interpreted as a single delimiter. For example, suppose that the text "8,,9" is parsed using the
Parse() function with \texttt{d} set to ",\). The Parse() function will treat the set of 3 consecutive commas as a single delimiter. Thus, only 2 tokens will be found.

**Parse() Code**

```cpp
inline std::vector<char*> Parse(char*s, //<===================================1D PARSER
    const char*d=" \	\n\f\r");//<----------------------------------DELIMITERS
std::vector<char*>V; /*<-*/for(s=strtok(s,d);s;s=strtok(0,d))V.push_back(s);
return V;
}
```

**Parse() Parameters**

- \texttt{s} points to text that will be parsed.
- \texttt{d} points to a set of token-delimiting characters. By default, spaces, commas, tabs, line feeds, form feeds, and carriage returns are all treated as delimiters.

**Parse() Return Value**

The Parse() function returns a vector of pointers. Each pointer points to the beginning of a token. The tokens are stored in the character array that was originally pointed to by the input parameter \texttt{s}.

**Parse() Example**

The following example uses the Parse() function to parse a sample block of text.

```cpp
#include "y_io_2.h"
int main(){
    char s[]="1.0 2.0 3.0
    4.0 5.0 6.0
    7.0 8.0 9.0";
    printf("ORIGINAL TEXT:\n%s\n",s);
    std::vector<char*>A=yIo2::Parse(s);
    printf("TEXT PARSED USING Parse():\n");
    for(int i=0,m=A.size();i<m;++i)printf("%s",A[i],i==m-1?"\n":" , ");
}
```

**OUTPUT:**

```
ORIGINAL TEXT:
1.0 2.0 3.0
4.0 5.0 6.0
7.0 8.0 9.0

TEXT PARSED USING Parse():
1.0 , 2.0 , 3.0 , 4.0 , 5.0 , 6.0 , 7.0 , 8.0 , 9.0
```

### 5.2 Parse2D() Function

The Parse2D() function can be used to separate the text contained in a character array into tokens. Tokens are separated by 2 types of delimiting characters. The first set of delimiters separates tokens within a row of data. Common examples are spaces, commas, and tabs. The
second set of delimiters separates data rows. Common examples are line feeds, form feeds, and carriage returns.

The Parse2D() function works by searching for rows, which are separated by row-delimiting characters. When a row is found, the delimiting character that immediately follows the row is replaced by the NULL character. The Parse() (not 2D) function is then used to parse the row. Any set of consecutive delimiting characters acts as a single delimiter.

Parse2D() Code

```cpp
inline std::vector<std::vector<char*>> Parse2D(std::string& s, const char*d=" ,	", const char*e="\n\f\r"){std::vector<std::vector<char*>> V; char*b,*c=new char[strlen(d)+strlen(e)+1]; for(strcat(strcpy(c,d),e);*(b=s+strspn(s,c));V.push_back(Parse(b,d))) s=b+strcspn(b,e),!*s?s:(*s=0,++s); delete[]c; return V; //........the number of columns per row may differ from row to row
}
```

Parse2D() Parameters

- **s** points to text that will be parsed.
- **d** points to a set of token-delimiting characters. Default values are spaces, commas, and tabs.
- **e** points to a set of row-delimiting characters. Default values are line feeds, form feeds, and carriage returns.

Parse2D() Return Value

The Parse2D() function returns a vector of vectors of pointers. Each pointer points to the beginning of a token. The tokens are stored in the character array that was originally pointed to by the input parameter **s**.

Parse2D() Example

The following example uses the Parse2D() function to parse a sample block of text.

```cpp
#include "y_io_2.h" //............................yIo2,<cstdio>{printf},<vector>
int main(){//<================SIMPLE EXAMPLE FOR THE Parse2D() FUNCTION
    char s[]="1.0 2.0 3.0\n4.0 5.0 6.0\n7.0 8.0 9.0";
    printf("ORIGINAL TEXT:\n%s\n\n",s);
    std::vector<std::vector<char*>> B=yIo2::Parse2D(s);
    printf("TEXT PARSED USING Parse2D():\n");
    for(int i=0,m=B.size();i<m;++i)for(int j=0,n=B[i].size();j<n;++j)
        printf("%s",B[i][j],j==n-1?\n": " );
} //~~~~~~~~YAGENAUT@GMAIL.COM~~~~~~~~~~~~~~~~~~~~~~~~~LAST~UPDATED~20JUL2014~~~~~~
```
5.3 PreParse() Function

The PreParse() function is useful for situations where tokens contain delimiting characters, such as when filenames contain spaces. The PreParse() function searches for delimiting characters that are not “protected” and replaces them with user-specified characters. Characters are assumed to be protected if they are encompassed by some sort of protecting characters (typically quotation marks, parentheses, brackets, or braces).

PreParse() Code

```c
inline char*PreParse(//<=======SELECTIVELY REPLACE DELIMITERS WITH A CHARACTER
    char*s,//<-----------------------------------------------A CHARACTER ARRAY
    const char*b="\",//<----------BEGINNING OF IGNORED-DELIMITERS SECTION
    const char*e="\",//<--------END OF IGNORED-DELIMITERS SECTION
    const char*d=" ,\t\n\f\r",//<----------------DELIMITING CHARACTERS
    char r='#'){//<------------------------REPLACEMENT CHARACTER
    const char*u;
    for(char*t=s,c=0;*t;++t){
        if(strchr(d,*t)&&!c)*t=r;
        else if(strchr(b,*t))u=b,b=e,e=u,c&1?++c:--c;
        else if(strchr(e,*t))u=b,b=e,e=u,c&1?--c:++c;}
    return s;
};//~~~~YAGENAUT@GMAIL.COM~~~~~~~~~~~~~~~~~~~~~~~~~LAST~UPDATED~20JUL2014~~~~~~
```

PreParse() Parameters

- **s** points to the text that will have its nonprotected delimiting characters replaced.
- **b** points to a list of characters that mark the beginning of a protected section of text. The default value is the double-quote character.
- **e** points to a list of characters that mark the end of a protected section of text. The default value is the double-quote character.
- **d** points to a set of delimiting characters.
- **r** specifies the character that will be used to replace nonprotected delimiting characters.
PreParse() Return Value

The PreParse() function returns the input pointer s.

PreParse() Example

The following example uses the PreParse() function to prepare a sample block of text for parsing. The text is then parsed using the Parse() function.

```cpp
#include "y_io_2.h" //......................................yIo2,<cstdio>{printf}
int main() { //<========================SIMPLE EXAMPLE FOR THE PreParse() FUNCTION
    char s[]="Alexis, Beatrice, Charlotte, "Dejah Thoris", Eunice"
    printf("ORIGINAL TEXT:\n%s\n\n",s);
    printf("PRE-PARSED TEXT:\n%s\n\n",yIo2::PreParse(s,"",""," ,"));
    printf("PARSED TEXT:\n\n")
    std::vector<char*>S=yIo2::Parse(s,"#");
    for(int i=0,n=S.size();i<n;++i)printf("%s\n",S[i]);
} //~~~~~~YAGENAUT@GMAIL.COM~~~~~~~~~~~~~~~~~~~~~~~~~LAST~UPDATED~20JUL2014~~~~~~
```

OUTPUT:

**ORIGINAL TEXT:**

Alexis, Beatrice, Charlotte, "Dejah Thoris", Eunice

**PRE-PARSED TEXT:**

Alexis###Beatrice###Charlotte###"Dejah Thoris"###Eunice

**PARSED TEXT:**

Alexis
Beatrice
Charlotte
Dejah Thoris
Eunice

6. Summary

A summary sheet is provided at the end of this report. It presents the yIo2 namespace, which contains the ReadTextFile(), WriteTextFile(), RemoveLineComments(), RemoveBlockComments(), Parse(), Parse2D(), and PreParse() functions.
# yIo2 Summary

**ReadTextFile()**

ReadTextFile() is a function to read text from a file. It takes a filename as a parameter and returns a vector containing the lines of text from the file.

```cpp
const std::vector<std::vector<char>> yIo2::ReadTextFile(const char* filename) {    
    std::vector<std::vector<char>> lines;    
    std::ifstream file(filename);    
    char line[1024];    
    while (file.getline(line, 1024)) {        
        lines.push_back(std::vector<char>(line));    
    }    
    return lines;    
}
```

**WriteTextFile()**

WriteTextFile() is a function to write text to a file. It takes two parameters: `filename` and `text`. It writes each line of `text` to a file specified by `filename`.

```cpp
void yIo2::WriteTextFile(const char* filename, const std::vector<std::vector<char>>& text) {    
    std::ofstream file(filename);    
    for (const auto& line : text) {        
        file << line << std::endl;    
    }    
}
```

**PreParse()**

PreParse() is a function that prepares the input text for further processing. It takes a character vector and a character as parameters. It checks if the character is a delimiter or not and returns the index of the next delimiter or the end of the string if no delimiter is found.

```cpp
int yIo2::PreParse(const std::vector<char>& text, char delimiter) {    
    int index = -1;    
    for (size_t i = 0; i < text.size(); ++i) {        
        if (text[i] == delimiter && index == -1) {            
            index = static_cast<int>(i);        
        }        
    }    
    return index;    
}
```

**Parse()**

Parse() is a function that parses a character vector. It takes a character vector as a parameter and returns a character vector that contains only non-delimiter characters.

```cpp
const std::vector<char> yIo2::Parse(const std::vector<char>& text) {    
    std::vector<char> result;    
    int start = 0, end = 0;    
    for (size_t i = 0; i < text.size(); ++i) {        
        if (text[i] != ' ') {            
            end = i;            
            result.push_back(text[start]);            
            start = end;        
        }        
    }    
    result.push_back(text[start]);    
    return result;    
}
```

**Parse2D()**

Parse2D() is a function that parses a two-dimensional character vector. It takes a character vector as a parameter and returns a two-dimensional character vector that contains only non-delimiter characters.

```cpp
const std::vector<std::vector<char>> yIo2::Parse2D(const std::vector<char>& text) {    
    std::vector<std::vector<char>> result;    
    int start = 0, end = 0;    
    for (size_t i = 0; i < text.size(); ++i) {        
        if (text[i] != ' ') {            
            end = i;            
            result.push_back(std::vector<char>(text.begin() + start, text.begin() + end));            
            start = end;        
        }        
    }    
    result.push_back(text.begin() + start);    
    return result;    
}
```

**yIo2 Example**

Here is an example of how to use the yIo2 library:

```cpp
#include "yIo2.h"    
#include <vector>    
#include <iostream>    

int main() {    
    std::vector<std::vector<double>> data = {        
        {1.0, 2.0, 3.0},        
        {4.0, 5.0, 6.0},        
        {7.0, 8.0, 9.0}    
    };    
    yIo2::Parse2D(data);    
    return 0;    
}
```

**yIo2 Summary**

Original Text:

```
Hello/*sample comment*/ World!
```

Output:

```
Hello World!
```

**Parse() Example**

Include "yIo2.h"; yIo2::Parse("The quick brown fox jumps over the lazy dog");

Output:

```
The quick brown fox jumps over the lazy dog
```

**Parse2D() Example**

Include "yIo2.h"; yIo2::Parse2D(("1 2 3
4 5 6
7 8 9");

Output:

```
1 2 3
4 5 6
7 8 9
```

**PreParse() Example**

Include "yIo2.h"; yIo2::PreParse(('z', 3));

Output:

```
3
```

**Original Text**

```
Hello/*sample comment*/ World!
```

**Output**

```
Hello World!
```

**Pre-Parse Text**

```
Alexis Beatrice Charlotte Dejah Thoris Eunice
```

**Parsed Text**

```
Alexis Beatrice Charlotte Dejah Thoris Eunice
```
7. References
