# 28.7 — Random file I/O



### The file pointer

Each file stream class contains a file pointer that is used to keep track of the current read/write position within the file. When something is read from or written to a file, the reading/writing happens at the file pointer's current location. By default, when opening a file for reading or writing, the file pointer is set to the beginning of the file. However, if a file is opened in append mode, the file pointer is moved to the end of the file, so that writing does not overwrite any of the current contents of the file.

### Random file access with seekg() and seekp()

So far, all of the file access we've done has been sequential -- that is, we've read or written the file contents in order. However, it is also possible to do random file access -- that is, skip around to various points in the file to read its contents. This can be useful when your file is full of records, and you wish to retrieve a specific record. Rather than reading all of the records until you get to the one you want, you can skip directly to the record you wish to retrieve.

Random file access is done by manipulating the file pointer using either seekg() function (for input) and seekp() function (for output). In case you are wondering, the g stands for "get" and the p for "put". For some types of streams, seekg() (changing the read position) and seekp() (changing the write position) operate independently -- however, with file streams, the read and write position are always identical, so seekg and seekp can be used interchangeably.

The seekg() and seekp() functions take two parameters. The first parameter is an offset that determines how many bytes to move the file pointer. The second parameter is an ios flag that specifies what the offset parameter should be offset from.

| Ios seek flag | Meaning |
|---|---|
| beg | The offset is relative to the beginning of the file (default) |
| cur | The offset is relative to the current location of the file pointer |
| end | The offset is relative to the end of the file |

A positive offset means move the file pointer towards the end of the file, whereas a negative offset means move the file pointer towards the beginning of the file.

Here are some examples:

```
inf.seekg(14, std::ios::cur); // move forward 14 bytes
inf.seekg(-18, std::ios::cur); // move backwards 18 bytes
inf.seekg(22, std::ios::beg); // move to 22nd byte in file
inf.seekg(24); // move to 24th byte in file
inf.seekg(-28, std::ios::end); // move to the 28th byte before end of the file
```

Moving to the beginning or end of the file is easy:

```
inf.seekg(0, std::ios::beg); // move to beginning of file
inf.seekg(0, std::ios::end); // move to end of file
```

Warning

In a text file, seeking to a position other than the beginning of the file may result in unexpected behavior.

In programming, a newline (‘\n’) is actually an abstraction.

- On Windows, a newline is represented as sequential CR (carriage return) and LF (line feed) characters (thus taking 2 bytes of storage).
- On Unix, a newline is represented as a LF (line feed) character (thus taking 1 byte of storage).

Seeking past a newline in either direction takes a variable number of bytes depending on how the file was encoded, which means results will vary depending on which encoding is used.

Also on some operating systems, files may be padded with trailing zero bytes (bytes that have value 0). Seeking to the end of the file (or an offset from the end of the file) will produce different results on such files.

Just to give you an idea of how they work, let’s do an example using seekg() and the input file we created in the last lesson. That input file looks like this:

```
This is line 1
This is line 2
This is line 3
This is line 4
```

Here is the example:

```
#include <fstream>
#include <iostream>
#include <string>

int main()
{
    std::ifstream inf{ "Sample.txt" };

    // If we couldn't open the input file stream for reading
    if (!inf)
    {
        // Print an error and exit
        std::cerr << "Uh oh, Sample.txt could not be opened for reading!\n";
        return 1;
    }

    std::string strData;

    inf.seekg(5); // move to 5th character
    // Get the rest of the line and print it, moving to line 2
    std::getline(inf, strData);
    std::cout << strData << '\n';

    inf.seekg(8, std::ios::cur); // move 8 more bytes into file
    // Get rest of the line and print it
    std::getline(inf, strData);
    std::cout << strData << '\n';

    inf.seekg(-14, std::ios::end); // move 14 bytes before end of file
    // Get rest of the line and print it
    std::getline(inf, strData); // undefined behavior
    std::cout << strData << '\n';

    return 0;
}
```

This produces the result:

```
is line 1
line 2
This is line 4
```

You may get a different result for the third line depending on how your file is encoded.

seekg() and seekp() are better used on binary files. You can open the above file in binary mode via:

```
std::ifstream inf {"Sample.txt", std::ifstream::binary};
```

Two other useful functions are tellg() and tellp(), which return the absolute position of the file pointer. This can be used to determine the size of a file:

```
std::ifstream inf {"Sample.txt"};
inf.seekg(0, std::ios::end); // move to end of file
std::cout << inf.tellg();
```

On the author's machine, this prints:

```
64
```

which is how long sample.txt is in bytes (assuming a newline after the last line).

Author's note

The result of `64` in the prior example occurred on Windows. If you run the example on Unix, you'll get `60` instead, due to the smaller newline representation. You may get something else if your file is padded with trailing zero bytes.

**Reading and writing a file at the same time using fstream**

The fstream class is capable of both reading and writing a file at the same time -- almost! The big caveat here is that it is not possible to switch between reading and writing arbitrarily. Once a read or write has taken place, the only way to switch between the two is to perform an operation that modifies the file position (e.g. a seek). If you don't actually want to move the file pointer (because it's already in the spot you want), you can always seek to the current position:

```
// assume iofile is an object of type fstream
iofile.seekg(iofile.tellg(), std::ios::beg); // seek to current file position
```

If you do not do this, any number of strange and bizarre things may occur.

(Note: Although it may seem that `iofile.seekg(0, std::ios::cur)` would also work, it appears some compilers may optimize this away).

One other bit of trickiness: Unlike ifstream, where we could say `while (inf)` to determine if there was more to read, this will not work with fstream.

Let's do a file I/O example using fstream. We're going to write a program that opens a file, reads its contents, and changes any vowels it finds to a '#' symbol.

```
#include <fstream>
#include <iostream>
#include <string>

int main()
{
    // Note we have to specify both in and out because we're using fstream
    std::fstream iofile{ "Sample.txt", std::ios::in | std::ios::out };

    // If we couldn't open iofile, print an error
    if (!iofile)
    {
        // Print an error and exit
        std::cerr << "Uh oh, Sample.txt could not be opened!\n";
        return 1;
    }

    char chChar{}; // we're going to do this character by character

    // While there's still data to process
    while (iofile.get(chChar))
    {
        switch (chChar)
        {
            // If we find a vowel
            case 'a':
            case 'e':
            case 'i':
            case 'o':
            case 'u':
            case 'A':
            case 'E':
            case 'I':
            case 'O':
            case 'U':

                // Back up one character
                iofile.seekg(-1, std::ios::cur);

                // Because we did a seek, we can now safely do a write, so
                // let's write a # over the vowel
                iofile << '#';

                // Now we want to go back to read mode so the next call
                // to get() will perform correctly.  We'll seekg() to the current
                // location because we don't want to move the file pointer.
                iofile.seekg(iofile.tellg(), std::ios::beg);

                break;
        }
    }
```

```
    return 0;
}
```

After running the above program, our Sample.txt file will look like this:

```
Th#s #s l#n# 1
Th#s #s l#n# 2
Th#s #s l#n# 3
Th#s #s l#n# 4
```

### Other useful file functions

To delete a file, simply use the remove() function.

Also, the is_open() function will return true if the stream is currently open, and false otherwise.

### A warning about writing pointers to disk

While streaming variables to a file is quite easy, things become more complicated when you're dealing with pointers. Remember that a pointer simply holds the address of the variable it is pointing to. Although it is possible to read and write addresses to disk, it is extremely dangerous to do so. This is because a variable's address may differ from execution to execution. Consequently, although a variable may have lived at address 0x0012FF7C when you wrote that address to disk, it may not live there any more when you read that address back in!

For example, let's say you had an integer named nValue that lived at address 0x0012FF7C. You assigned nValue the value 5. You also declared a pointer named *pnValue that points to nValue. pnValue holds nValue's address of 0x0012FF7C. You want to save these for later, so you write the value 5 and the address 0x0012FF7C to disk.

A few weeks later, you run the program again and read these values back from disk. You read the value 5 into another variable named nValue, which lives at 0x0012FF78. You read the address 0x0012FF7C into a new pointer named *pnValue. Because pnValue now points to 0x0012FF7C when the nValue lives at 0x0012FF78, pnValue is no longer pointing to nValue, and trying to access pnValue will lead you into trouble.

Warning

Do not write memory addresses to files. The variables that were originally at those addresses may be at different addresses when you read their values back in from disk, and the addresses will be invalid.