Subject: Filtering streams for bz2
Posted by Novo on Wed, 24 Jan 2018 19:28:53 GMT
View Forum Message <> Reply to Message

I've implemented bz2::CompressStream and bz2::DecompressStream. Source code is attached. Could you please add them to the Upp code?

UPD: My DecompressStream can read multistream bz2 files like this one

By default it will read all streams, but it would be nice to be able to allow DecompressStream/InFilterStream to read just one logical stream. The problem is that Filter is an Event (unlike More) and it doesn't allow to pass info about "end of stream" from Filter to InFilterStream the way More does that.

File Attachments
1) bz2.zip, downloaded 380 times

Subject: Re: Filtering streams for bz2
Posted by mirek on Sun, 28 Jan 2018 09:54:26 GMT
View Forum Message <> Reply to Message

Thanks, in trunk now.

Subject: Re: Filtering streams for bz2
Posted by Novo on Tue, 30 Jan 2018 01:54:23 GMT
View Forum Message <> Reply to Message

Thank you!

I've attached another patch which slightly changes InFilterStream and replaces type of InFilterStream::Filter (as I wrote in my previous message).

Could you please apply it as well?

File Attachments
1) bz2.02.zip, downloaded 462 times

Subject: Re: Filtering streams for bz2
Posted by mirek on Tue, 30 Jan 2018 10:46:23 GMT
View Forum Message <> Reply to Message

I am not particulary happy about changing the interface, but I can see the reason to signal eof in InFilterStream.

However, I am completely puzzled why have you changed OutFilterStream. AFAIK, all those new bool's there are ignored?

Mirek

Subject: Re: Filtering streams for bz2
Posted by Novo on Tue, 30 Jan 2018 13:24:39 GMT
View Forum Message <> Reply to Message

mirek wrote on Tue, 30 January 2018 05:46
However, I am completely puzzled why have you changed OutFilterStream. AFAIK, all those new bool's there are ignored?

Code wouldn't compile otherwise. I'm getting an error "void function 'Execute' should not return a value" (Clang) in the method OutFilterStream::Set "Filter = callback<F, F, const void *, int>(&filter, &F::Put);"

"bool Put(const void *ptr, int size)" is symmetric (used by both OutFilterStream and InFilterStream).
So, either interface of filter object should be redesigned (separate methods Put and Get), or OutFilterStream::Filter should be also be Gate.

Subject: Re: Filtering streams for bz2
Posted by mirek on Fri, 09 Feb 2018 08:54:26 GMT
View Forum Message <> Reply to Message

Novo wrote on Tue, 30 January 2018 14:24mirek wrote on Tue, 30 January 2018 05:46
However, I am completely puzzled why have you changed OutFilterStream. AFAIK, all those new bool's there are ignored?

Code wouldn't compile otherwise. I'm getting an error "void function 'Execute' should not return a value" (Clang) in the method OutFilterStream::Set "Filter = callback<F, F, const void *, int>(&filter, &F::Put);"

"bool Put(const void *ptr, int size)" is symmetric (used by both OutFilterStream and InFilterStream).
So, either interface of filter object should be redesigned (separate methods Put and Get), or OutFilterStream::Filter should be also be Gate.

Well, I really dislike changing so many things because of this.

I have tried to resolve the issue with Gate<> FilterEof in InFilterStream. That should work. Please check trunk.

BUT, after further thinking, I believe even that is overkill. In fact, I now believe that it should work without adding anything: After encountering BZ_STREAM_END, no more data are decompressed, so no more output is produced and we get nice eof via:

```
int InFilterStream::_Term()
{
 while(ptr == rdlim && !eof)
  Fetch();
 return ptr == rdlim ? -1 : *ptr;
}
```

(because after Fetch, ptr == rdlime)

What do you think?

Mirek

Subject: Re: Filtering streams for bz2
Posted by Novo on Sun, 11 Feb 2018 20:22:31 GMT
View Forum Message <> Reply to Message

mirek wrote on Fri, 09 February 2018 03:54

I have tried to resolve the issue with Gate<> FilterEof in InFilterStream. That should work. Please check trunk.

Solution with FilterEof works. Thank you.

mirek wrote on Fri, 09 February 2018 03:54
BUT, after further thinking, I believe even that is overkill. In fact, I now believe that it should work without adding anything: After encountering BZ_STREAM_END, no more data are decompressed, so no more output is produced and we get nice eof via:

```
int InFilterStream::_Term()
{
 while(ptr == rdlim && !eof)
  Fetch();
 return ptr == rdlim ? -1 : *ptr;
```

```
}
```

(because after Fetch, ptr == rdlime)

What do you think?

Mirek

Setting up of FilterEof cannot be skipped because in Fetch we have this:
```
 Stream::buffer = ptr = buffer.begin();
 rdlim = buffer.end();
```

And ptr != rdlim.

Below is my test.
```
#include <Core/Core.h>
#include <plugin/bz2/bz2.h>

using namespace Upp;

CONSOLE_APP_MAIN
{
 enum { BUF_SIZE = 65536 };
 const Vector<String>& cmdline = CommandLine();
 if (cmdline.GetCount() < 2)
  return;
 FileIn in(cmdline[0]);
 FileOut out(cmdline[1]);
#if 0
 bz2::DecompressStream bz2(in);
 char buff[BUF_SIZE];
 while (!bz2.IsEof()) {
  const int n = bz2.Get(buff, BUF_SIZE);
  out.Put(buff, n);
 }
#endif
#if 1
 // Read only a first stream.
 bz2::DecompressStream bz2(in, false);
 char buff[BUF_SIZE];
 while (!bz2.IsEof()) {
  const int n = bz2.Get(buff, BUF_SIZE);
  out.Put(buff, n);
 }
#endif
```

```
#if 0
 bz2::CompressStream bz2(out);
 char buff[BUF_SIZE];
 while (!in.IsEof()) {
  const int n = in.Get(buff, BUF_SIZE);
  bz2.Put(buff, n);
 }
#endif
}
```

Current version of code in svn/git works fine.
Thank you again!

Subject: Re: Filtering streams for bz2
Posted by mirek on Mon, 12 Feb 2018 01:10:22 GMT
View Forum Message <> Reply to Message

I believe: If Filter does not produce any output, then buffer.SetCount(0); holds and begin == end...
Actually, this also needs to work with partially filled buffer.

Anyway, can you post me an example .bz2 to test it with?

Subject: Re: Filtering streams for bz2
Posted by Novo on Mon, 12 Feb 2018 13:26:36 GMT
View Forum Message <> Reply to Message

mirek wrote on Sun, 11 February 2018 20:10
Anyway, can you post me an example .bz2 to test it with?
I'm using this file. It is quite big (14.5G). First stream has size 2.8K.
There is another file, which contains offsets of streams.

Subject: Re: Filtering streams for bz2
Posted by mirek on Thu, 15 Feb 2018 19:40:53 GMT
View Forum Message <> Reply to Message

OK, I have indentified the issue, but I am not sure that fixing that would work in all cases so I am staying with current solution for now.

(post edited to remove incorrect claims)

Mirek
