<http://mail-archives.apache.org/mod_mbox/lucene-dev/200110.mbox/%3CDB52A7625347D211A1D70060B06AB1E2035DC745@exchgcle.dev.netgenics.com%3E>

From dsnyder@netgenics.com Fri Oct 5 19:32:08 2001

Return-Path: <dsnyder@netgenics.com>

Mailing-List: contact lucene-dev-help@jakarta.apache.org; run by ezmlm

Delivered-To: mailing list lucene-dev@jakarta.apache.org

Delivered-To: moderator for lucene-dev@jakarta.apache.org

Received: (qmail 87776 invoked from network); 5 Oct 2001 19:32:08 -0000

Received: from exchgcle.netgenics.com (HELO exchgcle.dev.netgenics.com) (63.69.121.196)

by daedalus.apache.org with SMTP; 5 Oct 2001 19:32:08 -0000

Received: by exchgcle.dev.netgenics.com with Internet Mail Service (5.5.2653.19)

id <TJQ1M0FS>; Fri, 5 Oct 2001 15:42:42 -0400

Message-ID: <DB52A7625347D211A1D70060B06AB1E2035DC745@exchgcle.dev.netgenics.com>

From: "Snyder, David" <dsnyder@netgenics.com>

To: "'lucene-dev@jakarta.apache.org'" <lucene-dev@jakarta.apache.org>

Subject: RE: [Lucene-dev] Re: FW: Lucene 1.2 and directory write permissio

ns?

Date: Fri, 5 Oct 2001 15:42:41 -0400

MIME-Version: 1.0

X-Mailer: Internet Mail Service (5.5.2653.19)

Content-Type: multipart/alternative;

boundary="----\_=\_NextPart\_001\_01C14DD5.E5987030"

X-Spam-Rating: daedalus.apache.org 1.6.2 0/1000/N

This message is in MIME format. Since your mail reader does not understand

this format, some or all of this message may not be legible.

------\_=\_NextPart\_001\_01C14DD5.E5987030

Content-Type: text/plain;

charset="iso-8859-1"

Content-Transfer-Encoding: quoted-printable

I second that having the lock location configurable would be a "nice to

have"... but this brings up the problem of the searcher process being

configured differently than the updating process, which will defeat the

usefulness of the lock.

=20

I can imagine the sys-admin types wanting locks to go in a temp =

directory or

something like that, though.

=20

=20

Dave

=20

David Snyder=20

Se=F1or Software Engineer=20

NetGenics, Inc.=20

1717 E. 9th St., #1700=20

Cleveland, OH 44114=20

(216) 861-4007=20

-----Original Message-----

From: Dmitry Serebrennikov [mailto:dmitrys@earthlink.net]

Sent: Friday, October 05, 2001 3:24 PM

To: lucene-dev@jakarta.apache.org

Subject: [Lucene-dev] Re: FW: Lucene 1.2 and directory write =

permissions?

Just a though - as long as we are separating the locks, we should =

probably

make the location configurable and not require it to be a subdirectory =

of

the indexes. This will help in case the indexes come burned on a CD-ROM =

or

if the operating system does not support easy linking / permission =

magic.=20

Doug Cutting wrote:

Here's one vote for putting locks in a separate directory. Anyone =

dislike

that?

Doug

-----Original Message-----

From: Snyder, David [ mailto:dsnyder@netgenics.com

<mailto:dsnyder@netgenics.com> ]

Sent: Friday, October 05, 2001 11:23 AM

To: Doug Cutting

Subject: RE: Lucene 1.2 and directory write permissions?

The lock file synchronization is very handy for us as we do updates in =

a

separate process from searches... I was very pleased to see this in =

there!

I think splitting out the locks into a separate directory would solve =

our

problem... we have a reference set of data that we use for testing and =

our

sysadmin wants to make sure it doesn't accidentally get overwritten. =

It's

still possible (using links and other permission magic) but becomes =

more of

a maintenance headache. I definitely vote for the locks subdirectory =

idea.

It can be created when the index

es are initially loaded, we can go change

the permissions or maybe make it a link to a tmp directory or =

something, and

then our regular index files can be safe. Do you think this is =

something

very difficult to do? (I have yet to build Lucene myself, but I would =

love

to contribute... we are actually working on some XML based loaders that =

may

be of general interest)

Lucene has been great for us, but the way... we are indexing genetic =

data

(not the sequences themselves, but all the annotations and description =

stuff

that scientists tack on) and Lucene has been excellent... our indexes =

(we

use many with the multisearcher) are about 13 gigs now and Lucene has =

hardly

broken a sweat.

Thanks for your help,

Dave

David Snyder

Se=F1or Software Engineer

NetGenics, Inc.

1717 E. 9th St., #1700

Cleveland, OH 44114

(216) 861-4007

-----Original Message-----

From: Doug Cutting [ mailto:DCutting@grandcentral.com

<mailto:DCutting@grandcentral.com> ]

Sent: Friday, October 05, 2001 12:24 PM

To: 'Snyder, David'

Subject: RE: Lucene 1.2 and directory write permissions?

From: Snyder, David [ mailto:dsnyder@netgenics.com

<mailto:dsnyder@netgenics.com> ]

I've been porting our application to use the 1.2 release=20

candidate 1 build

and now have a problem opening searchers on our existing=20

indexes. I get a

Permission Denied exception... our permissions are set up to=20

allow reading

of the directory and contained files during a search, but not writing.

Hmm. That is a problem. The reader now creates a lock file while it =

is

opening the index to keep a writer process from deleting files while =

they're

being opened. When opening an index the reader must first read the =

list of

files to open, then open them. If between reading the list and opening =

a

file that file were to disappear, then the open would fail. This was =

the

longstanding race condition that is fixed by the lock files. A writing

process will now wait for the reader to open all of the files before

updating things.

Perhaps we should instead write lock files in a subdirectory of the =

index

named "locks". You could make that directory read/write, but make the

parent read-only. Alternately, we could have an flag that turns off =

the use

of lock files, for those who know that there is no other process that =

is

potentially simultaneously updating the index. Which approach would =

folks

prefer?

D

oug

------\_=\_NextPart\_001\_01C14DD5.E5987030

Content-Type: text/html;

charset="iso-8859-1"

Content-Transfer-Encoding: quoted-printable

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">

<HTML><HEAD>

<META HTTP-EQUIV=3D"Content-Type" CONTENT=3D"text/html; =

charset=3Diso-8859-1">

<META content=3D"MSHTML 6.00.2479.6" name=3DGENERATOR></HEAD>

<BODY>

<DIV><SPAN class=3D640052919-05102001><FONT face=3DArial =

color=3D#0000ff size=3D2>I=20

second that having the lock location configurable would be a "nice to=20

have"...&nbsp; but this brings up the problem of the searcher process =

being=20

configured differently than the updating process, which will defeat the =

usefulness of the lock.</FONT></SPAN></DIV>

<DIV><SPAN class=3D640052919-05102001><FONT face=3DArial =

color=3D#0000ff=20

size=3D2></FONT></SPAN>&nbsp;</DIV>

<DIV><SPAN class=3D640052919-05102001><FONT face=3DArial =

color=3D#0000ff size=3D2>I can=20

imagine the sys-admin types wanting locks to go in a temp directory or =

something=20

like that, though.</FONT></SPAN></DIV>

<DIV><SPAN class=3D640052919-05102001><FONT face=3DArial =

color=3D#0000ff=20

size=3D2></FONT></SPAN>&nbsp;</DIV>

<DIV><SPAN class=3D640052919-05102001><FONT face=3DArial =

color=3D#0000ff=20

size=3D2></FONT></SPAN>&nbsp;</DIV>

<DIV><SPAN class=3D640052919-05102001><FONT face=3DArial =

color=3D#0000ff=20

size=3D2>Dave</FONT></SPAN></DIV>

<DIV>&nbsp;</DIV>

<P><FONT face=3D"Times New Roman">David Snyder</FONT> <BR><I><FONT=20

face=3D"Times New Roman" size=3D2>Se=F1or Software Engineer</FONT></I> =

</P>

<P><FONT face=3D"Times New Roman" size=3D1>NetGenics, Inc.</FONT> =

<BR><FONT=20

face=3D"Times New Roman" size=3D1>1717 E. 9th St., #1700</FONT> =

<BR><FONT=20

face=3D"Times New Roman" size=3D1>Cleveland, OH 44114</FONT> <BR><FONT=20

face=3D"Times New Roman" size=3D1>(216) 861-4007</FONT> </P>

<BLOCKQUOTE>

<DIV class=3DOutlookMessageHeader dir=3Dltr align=3Dleft><FONT =

face=3DTahoma=20

size=3D2>-----Original Message-----<BR><B>From:</B> Dmitry =

Serebrennikov=20

[mailto:dmitrys@earthlink.net]<BR><B>Sent:</B> Friday, October 05, =

2001 3:24=20

PM<BR><B>To:</B> lucene-dev@jakarta.apache.org<BR><B>Subject:</B> =

[Lucene-dev]=20

Re: FW: Lucene 1.2 and directory write =

permissions?<BR><BR></FONT></DIV>Just a=20

though - as long as we are separating the locks, we should probably =

make the=20

location configurable and not require it to be a subdirectory of the =

indexes.=20

This will help in case the indexes come burned on a CD-ROM or if the =

operating=20

system does not support easy linking / permission magic. <BR><BR>Doug =

Cutting=20

wrote:<BR>

<BLOCKQUOTE=20

=

cite=3Dmid:4BC270C6AB8AD411AD0B00B0D0493DF0EE7C30@mail.grandcentral.com =

type=3D"cite"><PRE wrap=3D"">Here's one vote for putting locks in a =

separate directory. Anyone =

dislike<BR>that?<BR><BR>Doug<BR><BR>-----Original Message-----<BR>From: =

Snyder, David [<A class=3Dmoz-txt-link-freetext =

href=3D"mailto:dsnyder@netgenics.com">mailto:dsnyder@netgenics.com</A>]<=

BR>Sent: Friday, October 05, 2001 11:23 AM<BR>To: Doug =

Cutting<BR>Subject: RE: Lucene 1.2 and directory write =

permissions?<BR><BR><BR>The lock file synchronization is very handy for =

us as we do updates in a<BR>separate process from searches... I was =

very pleased to see this in there!<BR><BR>I think splitting out the =

locks into a separate directory would solve our<BR>problem... we have =

a reference set of data that we use for testing and our<BR>sysadmin =

wants to make sure it doesn't accidentally get overwritten. =

It's<BR>still possible (using links and other permission magic) but =

becomes more of<BR>a maintenance headache. I definitely vote for the =

locks subdirectory idea.<BR>It can be created when the index

es are initially loaded, we can go change<BR>the permissions or maybe =

make it a link to a tmp directory or something, and<BR>then our regular =

index files can be safe. Do you think this is something<BR>very =

difficult to do? (I have yet to build Lucene myself, but I would =

love<BR>to contribute... we are actually working on some XML based =

loaders that may<BR>be of general interest)<BR><BR> Lucene has been =

great for us, but the way... we are indexing genetic data<BR>(not the =

sequences themselves, but all the annotations and description =

stuff<BR>that scientists tack on) and Lucene has been excellent... our =

indexes (we<BR>use many with the multisearcher) are about 13 gigs now =

and Lucene has hardly<BR>broken a sweat.<BR><BR>Thanks for your =

help,<BR>Dave<BR><BR>David Snyder<BR>Se=F1or Software =

Engineer<BR><BR>NetGenics, Inc.<BR>1717 E. 9th St., #1700<BR>Cleveland, =

OH 44114<BR>(216) 861-4007<BR><BR><BR>-----Original =

Message-----<BR>From: Doug Cutting [<A =

class=3D"moz-txt-link-fre&#13;&#10;etext" =

href=3D"mailto:DCutting@grandcentral.com">mailto:DCutting@grandcentral.c=

om</A>]<BR>Sent: Friday, October 05, 2001 12:24 PM<BR>To: 'Snyder, =

David'<BR>Subject: RE: Lucene 1.2 and directory write =

permissions?<BR><BR><BR></PRE>

<BLOCKQUOTE type=3D"cite"><PRE wrap=3D"">From: Snyder, David [<A =

class=3Dmoz-txt-link-freetext =

href=3D"mailto:dsnyder@netgenics.com">mailto:dsnyder@netgenics.com</A>]<=

BR><BR>I've been porting our application to use the 1.2 release =

<BR>candidate 1 build<BR>and now have a problem opening searchers on =

our existing <BR>indexes. I get a<BR>Permission Denied exception... =

our permissions are set up to <BR>allow reading<BR>of the directory and =

contained files during a search, but not =

writing.<BR></PRE></BLOCKQUOTE><PRE wrap=3D""><!----><BR>Hmm. That is =

a problem. The reader now creates a lock file while it is<BR>opening =

the index to keep a writer process from deleting files while =

they're<BR>being opened. When opening an index the reader must first =

read the list of<BR>files to open, then open them. If between reading =

the list and opening a<BR>file that file were to disappear, then the =

open would fail. This was the<BR>longstanding race condition that is =

fixed by the lock files. A writing<BR>process will now wait for the =

reader to open all of the files before<BR>updating =

things.<BR><BR>Perhaps we should instead write lock files in a =

subdirectory of the index<BR>named "locks". You could make that =

directory read/write, but make the<BR>parent read-only. Alternately, =

we could have an flag that turns off the use<BR>of lock files, for =

those who know that there is no other process that is<BR>potentially =

simultaneously updating the index. Which approach would =

folks<BR>prefer?<BR><BR>D

oug<BR><BR></PRE></BLOCKQUOTE><BR></BLOCKQUOTE></BODY></HTML>

------\_=\_NextPart\_001\_01C14DD5.E5987030--