http://mail-archives.apache.org/mod\_mbox/lucene-dev/200110.mbox/raw/%3C20011008043128.95853.qmail@web20903.mail.yahoo.com%3E

From davekkw@yahoo.com Mon Oct 8 04:31:17 2001

Return-Path: <davekkw@yahoo.com>

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

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

Received: (qmail 19665 invoked from network); 8 Oct 2001 04:31:17 -0000

Received: from web20903.mail.yahoo.com (216.136.226.225)

by daedalus.apache.org with SMTP; 8 Oct 2001 04:31:17 -0000

Message-ID: <20011008043128.95853.qmail@web20903.mail.yahoo.com>

Received: from [165.21.122.230] by web20903.mail.yahoo.com via HTTP; Sun, 07 Oct 2001 21:31:28 PDT

Date: Sun, 7 Oct 2001 21:31:28 -0700 (PDT)

From: Dave Kor <davekkw@yahoo.com>

Subject: Re: CachingDirectory contribution

To: lucene-dev@jakarta.apache.org

In-Reply-To: <000001c14f9d$81346940$0100a8c0@wolverine>

MIME-Version: 1.0

Content-Type: text/plain; charset=us-ascii

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

> A while back I wrote a CachingDirectory

> implementation for Lucene which

> allows for caching an index on a local machine other

> than the "root"

> machine. This can be very useful for handling heavy

> load (such as David

> Snyder's 13 GB index :-))

13GB is considered a light load for Lucene. I am

currently running a lucene demo on my old but trusty

pentium-120mhz laptop with a 9GB index. It takes

Lucene about 20 seconds to handle the very first

query, probably because it is loading the index into

memory. All subsequent queries are instantaneous.

Anyway, I'm very curious as to how your directory

caching code works. What does it cache? files?

previously read data? Have you measured the amount of

performance improvement that was gained using your

caching system? What was the index size you used to

measure performance improvement?

If it does improve performance for huge indexes, I'll

+1.

This leads me to yet another of my buring questions..

has anyone pushed Lucene to its limits yet? If so,

what are they? What happens when Lucene hit its limit?

Does it throw an exception? coredump?

> In addition to that I could also provide my

> OracleDirectory

> implementation which stores all index files into an

> Oracle database

> instead of a file system structure. I haven't done a

> SQLServerDirectory,

> but I'm willing to implement it as well :)

I assume you're using BLOBs to store the index files?

What are the advantages of using the Oracle directory

over just using the file system?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Do You Yahoo!?

NEW from Yahoo! GeoCities - quick and easy web site hosting, just $8.95/month.

http://geocities.yahoo.com/ps/info1