http://mail-archives.apache.org/mod\_mbox/lucene-dev/200110.mbox/%3C15297.61795.915668.528304@cabernet.nelson.monkey.org%3E

From nelson@monkey.org Mon Oct 8 18:33:05 2001

Return-Path: <nelson@monkey.org>

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

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

Received: (qmail 85322 invoked from network); 8 Oct 2001 18:33:05 -0000

Received: from adsl-63-197-150-84.dsl.snfc21.pacbell.net (HELO cabernet.nelson.monkey.org) (foobar@63.197.150.84)

by daedalus.apache.org with SMTP; 8 Oct 2001 18:33:05 -0000

Received: by cabernet.nelson.monkey.org (Postfix, from userid 30193)

id E57018C00C; Mon, 8 Oct 2001 11:33:07 -0700 (PDT)

MIME-Version: 1.0

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

Content-Transfer-Encoding: 7bit

Message-ID: <15297.61795.915668.528304@cabernet.nelson.monkey.org>

Date: Mon, 8 Oct 2001 11:33:07 -0700

From: nelson@monkey.org (Nelson Minar)

To: lucene-dev@jakarta.apache.org

Subject: RE: CachingDirectory contribution

In-Reply-To: <4BC270C6AB8AD411AD0B00B0D0493DF0EE7C3C@mail.grandcentral.com>

References: <4BC270C6AB8AD411AD0B00B0D0493DF0EE7C3C@mail.grandcentral.com>

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

This is great discussion - thanks for explaining some of the limits, Doug.

I always wonder how this sort of email can turn into useful documentation.

>Performance should be more-or-less linear: a two-million document index will

>be almost twice as slow to search as a one-million document index.

Is performance linear in the number of documents? I would naively

think it would be linear in the number of terms. Also, how does

document size change things? Is searching 1 million 100k documents

twice as fast as 1 million 200k documents?

nelson@monkey.org

. . . . . . . . http://www.media.mit.edu/~nelson/