<http://mail-archives.apache.org/mod_mbox/lucene-dev/200110.mbox/%3C3BC0D8A5.4040308@earthlink.net%3E>

From dmitrys@earthlink.net Sun Oct 7 22:35:19 2001

Return-Path: <dmitrys@earthlink.net>

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

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

Received: (qmail 94326 invoked from network); 7 Oct 2001 22:35:19 -0000

Received: from relay4.uswest.net (HELO relay2.uswest.net) (63.226.138.12)

by daedalus.apache.org with SMTP; 7 Oct 2001 22:35:19 -0000

Received: (qmail 23196 invoked by uid 0); 7 Oct 2001 22:35:24 -0000

Received: from unknown (HELO earthlink.net) (65.100.117.194)

by relay2.uswest.net with SMTP; 7 Oct 2001 22:35:24 -0000

Message-ID: <3BC0D8A5.4040308@earthlink.net>

Date: Sun, 07 Oct 2001 16:35:17 -0600

From: Dmitry Serebrennikov <dmitrys@earthlink.net>

User-Agent: Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:0.9.2) Gecko/20010726 Netscape6/6.1

X-Accept-Language: en-us

MIME-Version: 1.0

To: lucene-dev@jakarta.apache.org

Subject: "optionally store document vectors in index"

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

Content-Transfer-Encoding: 7bit

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

Doug,

Could you give me some pointers to the approach you think should be

taken on this one? I'm in need of analyzing index terms from documents

retrieved by a query to identify further query expansion terms. We

already store the terms in a special field in the document, but that

leads to very slow processing because documents must be retrieved, field

re-parsed, etc. There are a couple of ideas that I'm going to try in the

next couple of weeks to fix this, but if you have a roadmap for this

already, I'd appreciate it. If the solution we end up with extends

Lucene, I'll be happy to contribute it back.

Thanks.

Dmitry.