http://mail-archives.apache.org/mod\_mbox/lucene-dev/200110.mbox/%3C3BC71FA3.6020109@earthlink.net%3E

From dmitrys@earthlink.net Fri Oct 12 16:52:00 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 49740 invoked from network); 12 Oct 2001 16:52:00 -0000

Received: from relay3.uswest.net (HELO relay1.uswest.net) (63.226.138.11)

by daedalus.apache.org with SMTP; 12 Oct 2001 16:52:00 -0000

Received: (qmail 71326 invoked by uid 0); 12 Oct 2001 16:52:03 -0000

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

by relay3.uswest.net with SMTP; 12 Oct 2001 16:52:03 -0000

Message-ID: <3BC71FA3.6020109@earthlink.net>

Date: Fri, 12 Oct 2001 10:51:47 -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: Re: Language recognition

References: <3BC6CB6F.EE8B4CB7@doris-frank.de>

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

That's great! I think our app may be getting ready for this in, say,

three to six months.

Frank Nestel wrote:

>Hi,

>

>this was a thread when lucene was still on Sourceforge.

>I've done a rough but working port of the text\_cat PERL

>script for n-gram based language guessing to Java. If this

>is useful, it can be found under

>

> http://frank.spieleck.de/ngram

>

>there are javadocs and a jar file. The source code

>is not yet available since I apparently need to reprogram

>a tiny but central class for copyright reasons. This is not

>difficult but I'm preparing for two weeks off right now and

>so it won't happen soon.

>

>For me it was just kind of an exercise, since I later

>realized that I still have a gap to make it work in my SMIDER

>project. If someone has a use for such a system, let me know

>so I can readjust this tasks priority for myself :-) Maybe

>one would even consider this a potential part of Lucene?

>Then I'd be glad to give that source code to apache.

>

>Regards,

>Frank

>