<http://mail-archives.apache.org/mod_mbox/lucene-dev/200110.mbox/%3C00ab01c14ced$a3ca7880$2400000a@trumpton.internal%3E>

Hi,

I've been lurking around the Lucene source code for about a week now...

There are a couple of things I can't work out how to do properly I'd be

grateful for any help with.

I'm having a bit of trouble using hit positions in a test application, the

results of which look like I may need to contribute some code to Lucene for

things to work as I'd like.

At the moment, I'm doing something along the lines of the following, to

retrieve hit positions:

// Open an index and retrieve the hit positions object

IndexReader reader = IndexReader.open("index\_file");

TermPositions hitPoints = reader.termPositions(new Term("contents",

"metal"));

TermDocs docs = (TermDocs) hitPoints;

// While a document remains, loop

while ( docs.next())

{

out.print("Finding hit values for document <b>"+ docs.doc()+"</b>");

for (int j=0; j<docs.freq(); j++)

{

// Output the hit position

out.print(", "+hitPoints.nextPosition());

}

out.println("<br>");

}

reader.close();

I'm not able to do a great deal with that information at the moment. What

I'd really like to be able to do is get the relevant info in my actual

search results loop. So I'd call something like this:

while (search\_results\_remain) {

Document doc = hits.doc(i);

int[] documentHitPositions = doc.getHitPositions();

// display fragments with 3 hits in the context text

String someContextInfo = hits.getContextInfo(i, 3);

}

My main difficulties with the existing way of doing things is:

1) The call to termPositions() doesn't integrate with QueryParser.parse()

and that appears to be the only correct way to use complex queries such as

wildcards, booleans, etc.

Is there any way, given a query, to get the list of 'Term' objects that were

created for the query? This would help me to an extent as I'd be able to

generate complete hit positions, rather than just for an arbitrary term.

2) Retrieving the hit positions doesn't integrate with the 'Hits' or

Document objects, where it would be the most convenient, imho, (as in my

example, above). Is it feasible to integrate such functionality?

Showing some amount of context for each search result is something that my

company considers to be really important for adopting any search engine.

Could anyone point me in the right direction for what changes, if any need

to be made to facilitate such a thing? If so, I may well be allowed to

contribute to Lucene on company time. From browsing the source and the

documentation, it appears that various things are in place to facilitate

implementing context information, I'm just not sure where exactly to

start...

Regards,

Lee Mallabone

Granta Design Ltd.