TermVectorComponent

Introduction

⚠ Solr 1.4 ⚠

The Term Vector Component (TVC) is a SearchComponent designed to return information about documents that is stored when setting the termVector attribute on a field:

```
<field name="features" type="text" indexed="true" stored="true" multiValued="true" termVectors="true"
termPositions="true" termOffsets="true"/>
```

For each document, the TVC can return the term vector, the term frequency, inverse document frequency, and position and offset information. As with most components, there are a number of options that are outlined in the samples below.

Sample Usage

All examples are based on using the Solr example server.

Enabling the TVC

Changes required in solrconfig.xml

You need to enable the TermVectorComponent in your solr configuration (this is already in the example solrconfig.xml):

```
<searchComponent name="tvComponent" class="org.apache.solr.handler.component.TermVectorComponent"/>
```

A RequestHandler configuration using this component could look like this:

```
<requestHandler name="tvrh" class="org.apache.solr.handler.component.SearchHandler">
        <lst name="defaults">
                <bool name="tv">true</bool>
        </lst>
        <arr name="last-components">
                <str>tvComponent</str>
        </arr>
</requestHandler>
```

Example Requests

In the example schema, the "includes" field has term vectors enabled. The following example HTTP request asks for the term vectors of all documents with something in the includes field.

[http://localhost:8983/solr/select/?qt=tvrh&q=includes:[*TO*]&fl=id]

In the example server, the component is associated with a request handler named tvrh, but you can associate it with any RequestHandler. To turn on the component for a request, add the `tv=true` parameter (or add it to your RequestHandler defaults configuration).

Example output: See TermVectorComponentExampleEnabled.

Options

- tv.tf - Return document term frequency info per term in the document.
- tv.df - Return the Document Frequency (DF) of the term in the collection. This can be expensive.
- tv.positions - Return position information.
- tv.offsets - Return offset information for each term in the document.
- tv.tf_idf - Calculates tf*idf for each term. Requires the parameters tv.tf and tv.df to be "true". This can be expensive. (not shown in example output)
- tv.all - If true, turn on extra information (tv.tf, tv.df, etc)
- tv.fl - (Solr3.1) Provides the list of fields to get term vectors for (defaults to fl)
- tv.docIds - List of Lucene document ids (not the Solr Unique Key) to get term vectors for.

An example HTTP request using these options:

[http://localhost:8983/solr/select/?qt=tvrh&q=includes:[*TO*]&fl=id&tv.all=true]

Per Field Options

([Solr3.1](#)) Options may be specified per-field, similar to the way per field options work in faceting, as in

- f.fieldName.tv.tf - Turns on Term Frequency for the fieldName specified.
- Similar for all the other options that are applicable to single fields

If you specify f.fieldName you must also explicitly declare &tv.fl or &fl

Example:

In this example, all features are requested, but then term frequency is turned off for the "includes" field (the only field returned)

```
http://localhost:8983/solr/select/?q=*%3A*&version=2.2&start=0&rows=100&indent=on&qt=tvrh&tv=true&tv.all=true&f.
includes.tv.tf=false&tv.fl=includes
```

Also:

In this example, all features are requested, but then offsets are turned off for the "includes" field (the only field returned)

```
http://localhost:8983/solr/select/?q=*%3A*&version=2.2&start=0&rows=100&indent=on&qt=tvrh&tv=true&tv.all=true&f.
includes.tv.offsets=false&tv.fl=includes
```

If you do not specify per field options but still specify a field, it will assume the general options.

Warnings

If a request field does not support the options specified, warnings will be returned indicating that the field does not support that option. There are three types of warnings:

1. noTermVector - The field does not store term vectors
2. noPositions - The field does not store positions
3. noOffsets - The field does not store offsets

Each of these items is a List of Strings containing the field name that does not support the option specified.

SolrJ client

There is a [patch](#) in progress for strongly-typed SolrJ support .