

# BloomIndexComponent

## Introduction

⚠ Solr4.0 - <https://issues.apache.org/jira/browse/SOLR-1375>

The BloomIndexComponent SearchComponent enables a fast and memory efficient membership test of an element in a read only set. False positives may be returned which can be tuned using a given probability (see the errorrate parameter of solrconfig.xml).

The Solr implementation generates the Bloom Filter per Lucene segment for a specified field by iterating over the term dictionary.

see [Wikipedia Bloom Filter](#) or [Hadoop BloomFilter](#)

## How it Works

To use the BloomIndexComponent, submit a query with the field and value you want to check the membership of. The host and check=true/false will be returned. If check=true then the segment it was found in will be returned along with the field and value.

The params are:

- q={String} - The id to lookup (i.e. id:1)

For a non-distributed membership test, the output returns true or false. For a distributed membership test, the output is the responses from the Solr shards.

## solrconfig.xml

```
<!-- Default Bloom Filter Implementation -->
<searchComponent name="bloom" class="org.apache.solr.handler.component.BloomIndexComponent">
  <!-- The following are all defaults -->
  <str name="dir">bloom</str> <!-- optional -->
  <str name="field">id</str> <!-- optional -->
  <boolean name="autoreload">true</boolean> <!-- optional -->
  <str name="classname">org.apache.solr.bloom.BloomKeySet</str> <!-- optional -->
  <float name="errorrate">0.003f</float> <!-- optional -->
  <int name="hashcount">5</int> <!-- optional -->
</searchComponent>

<requestHandler name="/bloom" class="org.apache.solr.handler.component.SearchHandler">
  <arr name="components">
    <str>bloom</str>
  </arr>
</requestHandler>
```

## Non-distributed Negative

```
http://localhost:8080/solr/bloom?q=id:1
```

Check the membership of the value 1 in the field id (which does not exist).

Results:

```
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">1</int>
  </lst>
  <str name="field">id</str>
  <str name="value">1</str>
  <str name="host">host.lightningstrike.com</str>
  <str name="exists">false</str>
</response>
```

## Non-distributed Positive

```
http://localhost:8080/solr/bloom?q=id:1
```

Check the membership of the value 1 in the field id (which does exist).

Results:

```
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">1</int>
  </lst>
  <str name="exists">true</str>
  <str name="segment">_0</str>
  <str name="field">id</str>
  <str name="value">1</str>
  <str name="host">host.lightningstrike.com</str>
</response>
```

## Distributed Positive

```
http://localhost:8080/solr/bloom?shards.qt=/bloom&q=id:1&shards=localhost:8080/solr,localhost:8081/solr
```

```
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">88</int>
  </lst>
  <lst name="host">
    <lst name="responseHeader">
      <int name="status">0</int>
      <int name="QTime">1</int>
    </lst>
    <str name="exists">true</str>
    <str name="segment">_3</str>
    <str name="field">id</str>
    <str name="value">1</str>
    <str name="host">host.lightningstrike.com</str>
  </lst>
  <str name="exists">true</str>
  <str name="distributed">true</str>
</response>
```

## Distributed Negative

```
http://localhost:8080/solr/bloom?shards.qt=/bloom&q=id:9999&shards=localhost:8080/solr,localhost:8081/solr
```

```
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">18</int>
  </lst>
  <str name="exists">false</str>
  <str name="distributed">true</str>
</response>
```

## Distributed Multiple Positive

http://localhost:8080/solr/bloom?shards.qt=/bloom&q=id:1&shards=localhost:8080/solr,localhost:8081/solr

```
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">20</int>
  </lst>
  <lst name="host">
    <lst name="responseHeader">
      <int name="status">0</int>
      <int name="QTime">0</int>
    </lst>
    <str name="exists">true</str>
    <str name="segment">_3</str>
    <str name="field">id</str>
    <str name="value">1</str>
    <str name="host">host.lightningstrike.com</str>
  </lst>
  <lst name="host">
    <lst name="responseHeader">
      <int name="status">0</int>
      <int name="QTime">0</int>
    </lst>
    <str name="exists">true</str>
    <str name="segment">_0</str>
    <str name="field">id</str>
    <str name="value">1</str>
    <str name="host">host2.lightningstrike.com</str>
    <str name="core"/>
  </lst>
  <str name="exists">true</str>
  <str name="distributed">true</str>
</response>
```