

# FieldCollapsingUncommitted

⚠ Solr4.0

- [Introduction](#)
- [Setup](#)
- [Request Parameters](#)
- [Examples](#)
- [Distributed field collapsing](#)
- [Other resources](#)

⚠ This page refers to functionality from [SOLR-236](#). It is not yet available in trunk.

## Introduction

"Used in order to collapse a group of results with similar value for a given field to a single entry in the result set. Site collapsing is a special case of this, where all results for a given web site is collapsed into one or two entries in the result set, typically with an associated "more documents from this site" link. See also Duplicate detection."

From [fast search](#) (TODO: this link is broken, fix it)

This topic was discussed a while ago: <http://www.nabble.com/result-grouping--tf2910425.html#a8131895>

## Setup

The easiest way to configure field collapsing is by overriding the query component. This can be achieved by adding the following xml in your solrconfig.xml:

```
<searchComponent name="query" class="org.apache.solr.handler.component.CollapseComponent" />
```

That is all, now you can have field collapse enabled searches. The [CollapseComponents](#) extends from the [QueryComponent](#), so a normal search is still possible.

If you wish to use both the [QueryComponent](#) and the [CollapseComponent](#) along side each other then you need to configure a little bit more in your solrconfig.xml. First, register the collapse searchComponent like this:

```
<searchComponent name="collapse" class="org.apache.solr.handler.component.CollapseComponent" />
```

Then reference that search component in a custom search handler. For example, you could modify the standard request handler to look like this:

```
<requestHandler name="standard" class="solr.SearchHandler" default="true">
  <!-- default values for query parameters -->
  <lst name="defaults">
    <str name="echoParams">explicit</str>

  </lst>
  <arr name="components">
    <str>collapse</str>
    <str>facet</str>
    <str>highlight</str>
    <str>debug</str>
  </arr>
</requestHandler>
```

Note that we have not included "query" in the list of component; the collapse handler implements query functionality itself.

In the latest patch it is possible to configure caching for the field collapsing execution. There are memory issues with this cache. Its therefore recommend to keep this cache small (e.g. with size 20) or to disable this cache. How big the cache should be depends on your environment.

This is an extra cache in addition to the already existing caches. It caches the result of the collapse logic and configured collapse collectors. The following xml configuration can be placed inside the solrconfig.xml as child of the config element.

```
<fieldCollapsing>

    <fieldCollapseCache
      class="solr.FastLRUCache"
      size="512"
      initialSize="512"
      autowarmCount="128" />

</fieldCollapsing>
```

If the field collapse cache is not configured then the field collapse logic will not be cached.

## Request Parameters

| param  | description  |  |  |
|--|--|--|--|
| collapse.type  | normal/adjacent – does this collapse all documents or just the ones that are next to each other. Defaults to normal  |  |  |
| collapse.field   | Which field to collapse. If this field is not specified then field collapsing is not enabled and falls back to the <a href="#">QueryComponent</a> to do a search.                |  |  |
| collapse.facet   | before/after – apply faceting before or after collapsing. Defaults to after  |  |  |
| collapse.max   | Deprecated use collapse.threshold instead. This parameter is removed in the latest patch.  |  |  |
| collapse.threshold   | The number of documents with the same value for collapse.field after which collapsing kicks in. The default value is one.  |  |  |
| collapse.maxdocs   | Maximum number of documents to process during field collapse. This parameter defaults to one greater than the largest document number.   |  |  |
| collapse.info.doc  | Return collapse count for each document? Defaults to true  |  |  |
| collapse.info.count  | Return collapse count for each field value? Defaults to true   |  |  |
| collapse.includeCollapsedDocs.fl   | Parameter indicating to return the collapsed documents in the response and what fields to return in comma separated manner. A value * indicates that all fields will be returned |  |  |
| collapse.debug   | wheter to include collapse debug information   |  |  |
| <ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="749267a4-e41a-415d-be76-1e2d126a561c"><ac:plain-text-body><![CDATA[ | collapse.aggregate   | Execute aggregate functions on the collapsed documents. The parameter expect the functions in the following format: function_name(field_name) [, function_name(field_name)]. So for example: sum(stock), avg(weight). Currently there are four functions available: min (...), max(...), sum(...), avg(...). The functionality is available from the patch added at 2009-10-25 10:13 PM. | ]]></ac:plain-text-body></ac:structured-macro> |

## Examples

Using the example data:

Collapse all documents using 'manu\_exact' and 'normal' collapse type: [http://localhost:8983/solr/select?q=&#42;.&#42;&collapse.field=manu\\_exact&collapse.threshold=1&collapse.type=normal](http://localhost:8983/solr/select?q=&#42;.&#42;&collapse.field=manu_exact&collapse.threshold=1&collapse.type=normal)

```

<lst name="collapse_counts">
  <str name="field">manu_exact</str>
  <lst name="results">
    <lst name="F8V7067-APL-KIT">
      <int name="collapseCount">1</int>
      <str name="fieldValue">Belkin</str>
    </lst>
    <lst name="TWINX2048-3200PRO">
      <int name="collapseCount">3</int>
      <str name="fieldValue">Corsair Microsystems Inc.</str>
    </lst>
    <lst name="VDBDB1A16">
      <int name="collapseCount">1</int>
      <str name="fieldValue">A-DATA Technology Inc.</str>
    </lst>
    <lst name="0579B002">
      <int name="collapseCount">1</int>
      <str name="fieldValue">Canon Inc.</str>
    </lst>
    <lst name="SOLR1000">
      <int name="collapseCount">1</int>
      <str name="fieldValue">Apache Software Foundation</str>
    </lst>
  </lst>
</lst>

```

Collapse all documents using 'manu\_exact' and 'adjacent' collapse type: [http://localhost:8983/solr/select?q=&#42;.&#42;&collapse.field=manu\\_exact&collapse.threshold=1&collapse.type=adjacent](http://localhost:8983/solr/select?q=&#42;.&#42;&collapse.field=manu_exact&collapse.threshold=1&collapse.type=adjacent)

```

<lst name="collapse_counts">
  <str name="field">manu_exact</str>
  <lst name="results">
    <lst name="F8V7067-APL-KIT">
      <int name="collapseCount">1</int>
      <str name="fieldValue">Belkin</str>
    </lst>
    <lst name="TWINX2048-3200PRO">
      <int name="collapseCount">1</int>
      <str name="fieldValue">Corsair Microsystems Inc.</str>
    </lst>
    <lst name="TWINX2048-3200PRO-payload">
      <int name="collapseCount">1</int>
      <str name="fieldValue">Corsair Microsystems Inc.</str>
    </lst>
  </lst>
</lst>

```

The response is centred around collapse groups. A collapse group represents documents that were collapsed during the search. A collapse group is identifier by the most relevant document of that collapse group, which is document that did not get collapsed and remained present in the search result. So the ids like 233238 are from documents that are also present in the search result.

## Distributed field collapsing

In a distributed environment fieldcollapsing is supported in a limited manner. While indexing you must make sure that the documents of a collapse group are not scattered across different shards. Documents of a collapse group must reside on the same shard, failing to do so will corrupt your search results. Doing a distributed search with collapsing requires not extra parameters to be send with the request. For example the following request is sufficient: [http://localhost:8080/solr/select?q=solr&collapse.field=my\\_field&shards=localhost:55527/solr,localhost:55529/solr](http://localhost:8080/solr/select?q=solr&collapse.field=my_field&shards=localhost:55527/solr,localhost:55529/solr)

## Other resources

Some other resources regarding to field collapsing:

- [Result grouping / field collapsing with Solr](#)

If anyone has links about this topic feel free to add it.