# Exchanges



## Exchanges in Nutch

An exchange is the component, which acts in indexing job and decides which index writer a document should be routed to. This component is based on plugins behavior and Nutch includes these exchanges out-of-the-box:

| Exchange | Description |
|---|---|
| exchange-jexl | Plugin of Exchange component based on JEXL expressions. |

## Structure of exchanges.xml

The exchanges to be used must be configured in the exchanges.xml file, included in the official Nutch distribution. The structure of this file consists mainly of a list of exchanges (`<exchanges>` element) and will be explained on this section:

```
<exchanges>
<exchange id="[exchange_id]" class="[implementation_class]">
<writers>
...
</writers>
<params>
...
</params>
</exchange>
...
</exchanges>
```

Each `<exchange>` element has two mandatory attributes:

1. `[exchange_id]` is a unique identification for each configuration. It is used by Nutch to distinguish each one, even when they are for the same exchange implementation and this ID allows to have multiple instances for the same exchange, but with different configurations.
2. `[implementation_class]` corresponds to the canonical name of the class that implements the Exchange extension point. For the exchanges provided by Nutch out-of-the-box, the possible values of `[implementation_class]` are:

| Exchange | Implementation class |
|---|---|
| exchange-jexl | org.apache.nutch.exchange.jexl.JexlExchange |

### Writers section

The `<writers>` element is independent for each configuration and contains a list of `<writer id="[id]">` elements, where `[id]` indicates the ID of index writer where the documents should be routed. See IndexWriters for more information about how to configure the index writers properly.

### Params section

The `<params>` element is where the parameters that the exchange needs are specified. Each parameter has the form `<param name="[name]" value="[value]"/>` and the values it can take depend on the exchange that you want to configure. Below is a description of the arguments of each exchange provided by Nutch out-of-the-box individually.

| Parameter name | Description | Default value |
|---|---|---|

| | | |
|---|---|---|
| expr | JEXL expression used to validate each document. The variable "doc" can be used on the expressions and represents the document itself. For example, the expression doc.getFieldValue('host')=='example.org' will match the documents where the "host" field has the value "example.org" | |

# Exchange behavior

The exchange component is in charge to route documents to the configured index writers, depending on whether documents match a piece of logic (defined for each exchange) or not. This component processes the documents one by one. If a document matches an exchange, then the document will be sent to the index writers declared in the exchange's configuration. If a document doesn't match any exchange, then it will be routed to the index writers indicated by the "default" exchange. If no exchange is configured, documents will be routed to all configured index writers.

## Default exchange

The "default" exchange is included into the core exchange component. So, you don't have to enable any plugin to use it. Its main functionality is to route the documents that don't match the other exchanges.

> ⚠ **Absence of default exchange**
>
> If the default exchange is not configured in the exchanges.xml file, but there are other exchanges, the documents that do not match will be discarded.

## Use case 1

There isn't any exchange configured (out-of-the-box behavior). So, the exchanges.xml file looks like:

```
<exchanges xmlns="http://lucene.apache.org/nutch"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://lucene.apache.org/nutch exchanges.xsd">
</exchanges>
```

**Result:** The documents will be routed to all configured index writers.

## Use case 2

We have two exchanges (jexl and default) and our exchanges.xml file looks like:

```
<exchanges xmlns="http://lucene.apache.org/nutch"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://lucene.apache.org/nutch exchanges.xsd">
<exchange id="exchange_jexl_1" class="org.apache.nutch.exchange.jexl.JexlExchange">
<writers>
<writer id="indexer_solr_1" />
<writer id="indexer_rabbit_1" />
</writers>
<params>
<param name="expr" value="doc.getFieldValue('host')=='example.org'" />
</params>
</exchange>

<exchange id="default" class="default">
<writers>
<writer id="indexer_dummy_1" />
</writers>
<params />
</exchange>
</exchanges>
```

We have 4 index writers properly configured in index-writers.xml file:

```
<writer id="indexer_solr_1" class="org.apache.nutch.indexwriter.solr.SolrIndexWriter">
...
</writer>
<writer id="indexer_solr_2" class="org.apache.nutch.indexwriter.solr.SolrIndexWriter">
...
</writer>
<writer id="indexer_rabbit_1" class="org.apache.nutch.indexwriter.rabbit.RabbitIndexWriter">
...
</writer>
<writer id="indexer_dummy_1" class="org.apache.nutch.indexwriter.dummy.DummyIndexWriter">
...
</writer>
```

**Result:** The documents which the value of "host" field is "example.org" will be sent to indexer_solr_1 and indexer_rabbit_1. The rest of documents where "host" is different to "example.org" do not match with exchange_jexl_1 exchange and will be sent where the default exchange says; in this case to indexer_dummy_1.



**indexer_solr_2 not used**

The index writer "indexer_solr_2" is not used. Which means that none of the documents will be routed to this index writer.