

Apache UIMA Solrcas documentation

Written and maintained by the Apache UIMA Development Community

Version 2.3.1

Copyright © 2006, 2011 The Apache Software Foundation

License and Disclaimer. The ASF licenses this documentation to you under the Apache License, Version 2.0 (the "License"); you may not use this documentation except in compliance with the License. You may obtain a copy of the License at <http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, this documentation and its contents are distributed under the License on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Trademarks. All terms mentioned in the text that are known to be trademarks or service marks have been appropriately capitalized. Use of such terms in this book should not be regarded as affecting the validity of the the trademark or service mark.

Publication date August, 2011

Table of Contents

Introduction	v
1. Configuration	1
2. The mapping file	3

Introduction

The Solr CAS Consumer (Solrcas) is responsible to write UIMA CAS objects to an Apache Solr instance.

It uses SolrJ client classes to execute local or remote updates to the specified Solr instance.

Chapter 1. Configuration

To use Solrcas the following parameters have to be specified:

- `mappingFile` : identifies where is the file which holds information about which (and how) UIMA objects must be sent to which Solr fields.
- `solrInstanceType` : this has to be `http`.
- `solrPath` : If the `solrInstance` value is 'http' this represents the URL to the remote Solr instance.

Chapter 2. The mapping file

The mapping file holds information about mapping between CAS properties, types and features and Solr fields.

Here is a solrMapping.xml sample:

```
<solrMapping>
  <documentText>text</documentText>
  <documentLanguage>language</documentLanguage>
  <fsMapping>
    <type name="uima.jcas.tcas.Annotation">
      <map feature="coveredText" field="annotation"/>
    </type>
  </fsMapping>
</solrMapping>
```

The *documentText* element holds the field name in which the `Cas.getDocumentText()` value will be indexed.

The *documentLanguage* element holds the field name in which the `Cas.getDocumentLanguage()` value will be indexed.

The *fsMapping* element will hold a list of *types*. For each *type* specified a *map* between a *feature* and a *field* will be defined. As the `getCoveredText()` of `Annotation` objects is not a `Feature` the `coveredText` feature name will be automatically associated with the `Annotation.getCoveredText()` value (just like a common feature).

In the sample above the `Cas.getDocumentText()` will be written inside the `text` field, the `Cas.getDocumentLanguage()` will be written inside the `language` field and the `Annotation.getCoveredText()` of each `uima.jcas.tcas.Annotation` object will be written inside an `annotation` field in Solr.

Note that `documentText` and `documentLanguage` are all optional.

