

**git\_comments:**

1. construct BreakIterator
2. \* \* Licensed to the Apache Software Foundation (ASF) under one or more \* contributor license agreements. See the NOTICE file distributed with \* this work for additional information regarding copyright ownership. \* The ASF licenses this file to You under the Apache License, Version 2.0 \* (the "License"); you may not use this file 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, software \* distributed under the License is distributed 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.
3. construct Locale
4. ////////////////////////////////////// SolrInfoMBeans methods //////////////////////////////////////
5. \* \* Licensed to the Apache Software Foundation (ASF) under one or more \* contributor license agreements. See the NOTICE file distributed with \* this work for additional information regarding copyright ownership. \* The ASF licenses this file to You under the Apache License, Version 2.0 \* (the "License"); you may not use this file 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, software \* distributed under the License is distributed 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.
6. ////////////////////////////////////// SolrInfoMBeans methods //////////////////////////////////////
7. \* \* Licensed to the Apache Software Foundation (ASF) under one or more \* contributor license agreements. See the NOTICE file distributed with \* this work for additional information regarding copyright ownership. \* The ASF licenses this file to You under the Apache License, Version 2.0 \* (the "License"); you may not use this file 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, software \* distributed under the License is distributed 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.
8. Load the BoundaryScanners
9. Thread safe registry
10. Make sure we loaded one fragListBuilder Make sure we loaded two fragmentsBuilders
11. Make sure we loaded two boundaryScanners
12. language and country are used when constructing Locale object. And the Locale object will be used when getting instance of BreakIterator
13. type should be one of CHARACTER, WORD(default), LINE and SENTENCE

**git\_commits:**

1. **summary:** SOLR-2749: add boundary scanner feature  
**message:** SOLR-2749: add boundary scanner feature git-svn-id: <https://svn.apache.org/repos/asf/lucene/dev/trunk@1170616> 13f79535-47bb-0310-9956-ffa450edef68

**github\_issues:****github\_issues\_comments:****github\_pulls:****github\_pulls\_comments:****github\_pulls\_reviews:****jira\_issues:**

1. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.
2. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.  
**label:** test
3. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.
4. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.
5. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.
6. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.
7. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.
8. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.
9. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.
10. **summary:** use BoundaryScanner in Solr FVH  
**description:** After LUCENE-1824 committed, Solr FragmentsBuilder can snip off at the "natural" boundary by nature. But to bring out the full feature, Solr should take care of arbitrary BoundaryScanner in solrconfig.xml.

**jira\_issues\_comments:**

1. Draft, halfway patch.
2. **body:** New patch. Almost done except test cases.  
**label:** test
3. New patch. I added test case. Will commit tonight.
4. trunk: Committed revision 1170616. 3x: Committed revision 1170620.
5. Bulk close after 3.5 is released
6. Hi, I just installed trunk, and I'm still seeing this problem. Not sure how to provide a STR, but in snippets I'm seeing words cut in the middle rather than at their boundaries. If it's useful, the query I'm sending is: INFO: [] webapp=/solr path=/select/ params={ sort=score+asc&

```
fl=id,absolute_url,court_id,local_path,source,download_url,status,dateFiled& hl.fl=text,caseName,westCite,docketNumber,lexisCite,court_citation_string&
f.text.hl.snippets=5& hl=true& q=willingness& fq=dateFiled:{*+TO+*}& fq={!tag%3Ddt}court_exact:
("ca5"+OR+"ca4"+OR+"ca7"+OR+"ca1"+OR+"ca3"+OR+"ca2"+OR+"scotus"+OR+"ca9"+OR+"ca8"+OR+"all"+OR+"ca11"+OR+"ca10"+OR+"cadc"+OR+"cafc"
fq={!tag%3Ddt}status_exact:("Non-Precedential"+OR+"Relating-
to+orders"+OR+"Precedential"+OR+"Errata")&f.docketNumber.hl.alternateField=docketNumber& f.docketNumber.hl.fragListBuilder=single&
f.lexisCite.hl.fragListBuilder=single& f.caseName.hl.fragListBuilder=single& f.westCite.hl.fragListBuilder=single&
f.court_citation_string.hl.fragListBuilder=single& f.text.hl.alternateField=text& f.caseName.hl.alternateField=caseName&
f.court_citation_string.hl.alternateField=court_citation_string f.lexisCite.hl.alternateField=lexisCite& f.westCite.hl.alternateField=westCite&
f.text.hl.maxAlternateFieldLength=500& } And I'm getting a snippet that contains: ...g and willingness to read with care." Rosenau v. Unifund Corp., 539 F.3d
218, 221 (3d Cir. 2008) (internal... You can see the first and last word are both cut off.
```

7. Are you sure that you have boundaryScanner tag in your solrconfig.xml? {code} <boundaryScanner name="default" default="true" class="solr.highlight.SimpleBoundaryScanner"> <lst name="defaults"> <str name="hl.bs.maxScan">10</str> <str name="hl.bs.chars">.,!? &#9;&#10;&#13;</str> </lst> </boundaryScanner> {code}
8. Oops. I suppose that would do it, huh? Is there a reason why this isn't in the config by default? Seems like one more place for a newbie to Solr (like myself) to miss a useful feature. I'll update the wiki as well to make it clear what the BoundaryScanners are for and what a valid entry for them looks like.