

AIDA

Training Solr use cases

Luís Faria <u>Ifaria@keep.pt</u> Miguel Guimarães <u>mguimaraes@keep.pt</u>



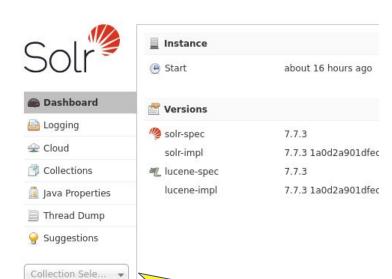
Apache Solr access

Solr can be accessed at: https://scala.meemoo.be/solr/

It is protected with a basic authentication layer

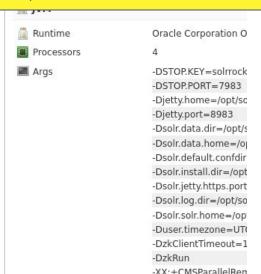
Password were provided by KEEP Solutions

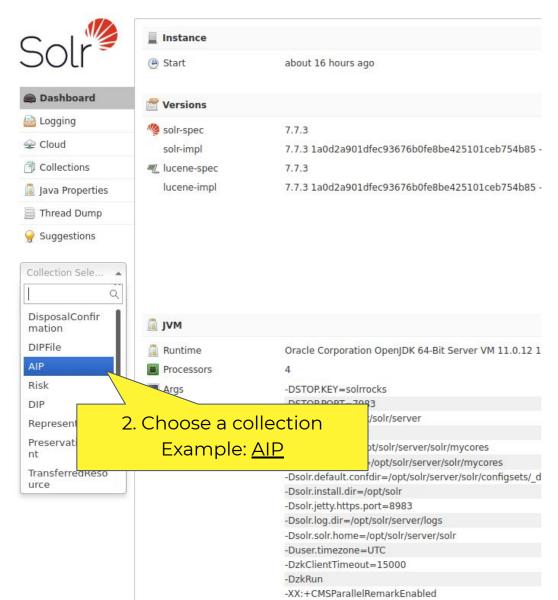
Documentation can be found at: <u>Apache Solr Reference Guide 7.7</u>

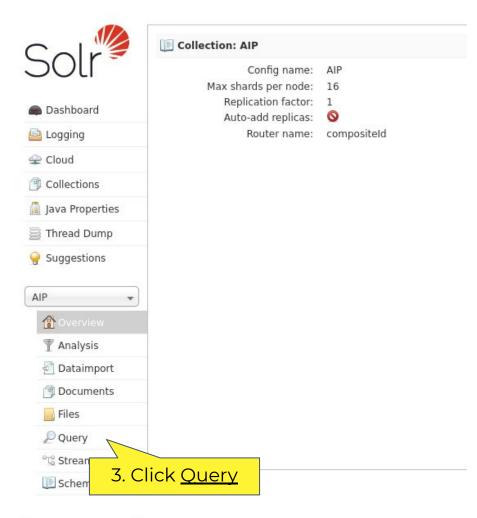


Core Selector

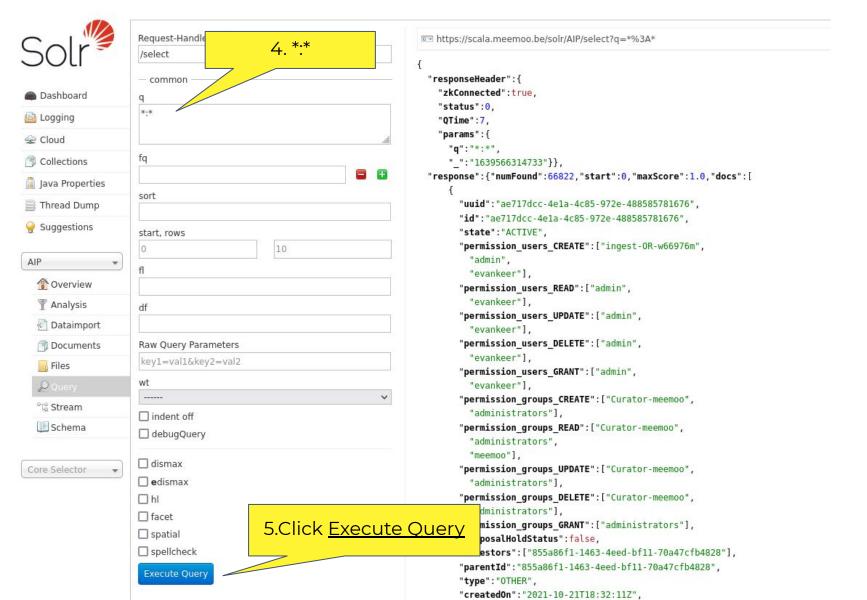
1. Open <u>Collection Sellector</u>















Collection: AIP

1. EAD metadata

- a. All text fields (search operator)
- b. All boolean or controlled vocabulary fields (dynamic fields, data types)

2. Meemoo.xml metadata

- a. Synchronisation AIP status
- b. Pruned / not pruned
- c. All text fields (search operator)



Collection: AIP

3. AIP.json metadata

- a. id
- b. parentId
- c. type
- d. state
- e. ingestJobId
- f. hasRepresentation
- g. createdOn datetime (before, after, on, between)
- h. updatedOn datetime (before, after, on, between)



Collection: File

1. PREMIS object metadata

- a. size (bigger than, smaller than, equals, between)
- b. formatDesignation
- c. formatVersion
- d. formatPronom: formatRegistry/formatRegistryKey
- e. formatMimetype: formatRegistry/formatRegistryKey
- f. originalName



Collection: Representation

1. AIP.json metadata

- a. id
- b. original (boolean facet)
- c. representationStates
- d. type
- e. createdOn datetime (before, after, on, between)
- f. updatedOn datetime (before, after, on, between)



Collection: PreservationEvent

1. PREMIS event metadata

- a. id
- b. eventType
- c. eventDateTime (before, after, on, between)
- d. eventOutcome
- e. linkingAgentIdentifierValue
- f. linkingObjectIdentifier (linkingSourceObjectIdentifier and linkingOutcomeObjectIdentifier)

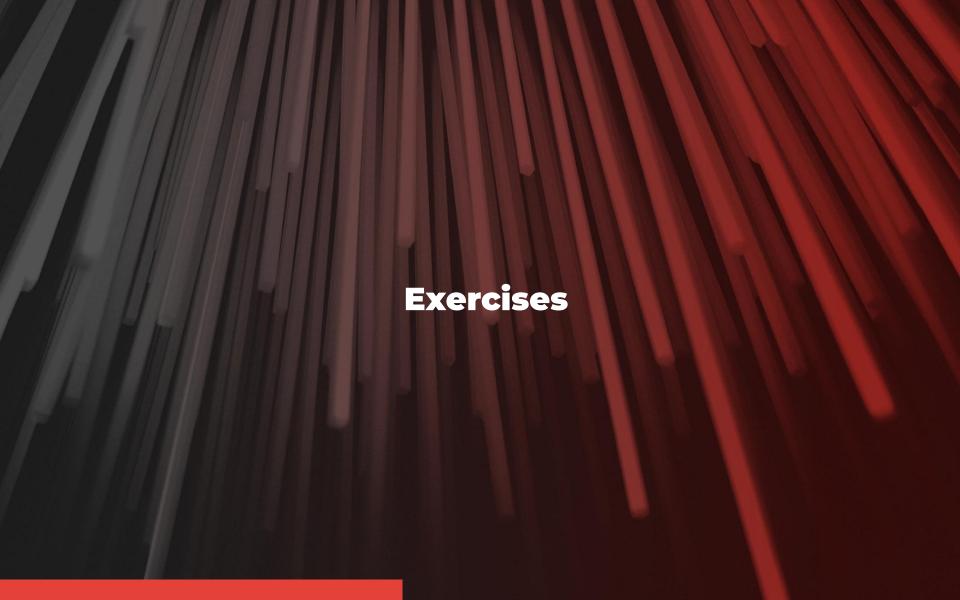




Advanced queries: Boolean operators

Operator	Description
AND	Requires both terms to be present for a match.
OR	Requires that either term (or both terms) be present for a match.
+	Requires that the following term be present.
-	Prohibits the following term (similar to what Google does)
NOT	Requires that the following term not be present.

Advanced queries Demonstration





For the collection **AIP**:

- 1. Search in all AIPs with string 'jkleevens'
- Show the facets for the field 'meemoo_sync_aip_status_txt'
- 3. How many AIPs are pruned?
- 4. Search the AIP with id '80ed7f6f-5e56-4c91-a9bd-5458e13154b3' and return the value of the field 'title'
- 5. How many AIPs with parentld 'de9eaa85-edb2-4bb3-aaca-cb3ec84a6e3f'?



Exercise 1 (continuation)

- 6. Show the facets for the field 'type'
- 7. How many AIPs are in the state 'ACTIVE'?
- 8. What is the ingestJobId that created the higher number of AIPs?
- 9. Search all AIPs with representations (hasRepresentations)
- 10. Search all AIPs created on December
- 11. Search all AIPs created between November 10 and now
- 12. Search all AIPs updated on '2021-12-08T15:32:24Z'



For the collection File:

- 1. Search all Files bigger than 9000 bytes
- 2. How many files are between 19968 and 20000 bytes (inclusive)?
- 3. How many files are between 19968 and 20000 bytes (exclusive)?
- 4. How many 'Portable Network Graphics 1.0' files are?
- 5. Facet the files by formatPronom
- **6**. Facet the files by formatMimetype



Exercise 2 (continuation)

- **7**. What is the MIME type with most files?
- **8**. Search the file with the originalName containing the keyword Heylighen. What is the aipld of the resulting query?



For the collection **Representation**:

- 1. Search all representations with id **'rep2'**
- 2. Facet the representations by original field
- 3. Face the representations by representationStates field and list its values
- 4. How many representations there are with the type 'MIXED'?
- 5. Search all representations created on October
- 6. Search all representations updated today



For the collection **PreservationEvent**:

- Search the preservation event with id
 'urn:roda:premis:event:76efb77d-la80-4ala-bc74-lb3906ef4b05'
- 2. Facet the preservation events by eventType field
- 3. How many preservation events there are of eventType 'message digest calculation'?
- 4. How many preservation events have the eventOutcome 'PARTIAL_SUCCESS'?



Exercise 4 (continuation)

For the collection **PreservationEvent**:

- 5. Search all preservation events with linkingAgentIdentifier
 'linkingAgentIdentifier:"{\"type\":\"URN\",\"value\":\"urn:roda:premis:agent:inge
 st-OR-w66976m\",\"roles\":[]}"'
- 6. Search all preservation events with linkingOutcomeObjectIdentifier
- 7. '{\"type\":\"URN\",\"value\":\"urn:roda:aip:00e3894e-ff3f-4cad-8ce9-afed74e50ca 8\",\"roles\":[\"outcome\"]}'



For the collection AIP:

1. Search all AIPs that are pruned and the parentld

'ed3964e1-d8a7-4786-a594-8f87c08a90f5'

2. Search all AIPs that descriptiveMetadataId does not contain the string meemoo.xml



For the collection File:

- 1. Search all files that contains the string Word in the formatDesignation and the size is bigger than 50000 (inclusive)
- 2. Search all files that contains the string 'pdf' in fileFormat however requiring to have an X



For the collection **PreservationEvent**:

- Search all preservation events with the eventOutcome 'SUCCESS' or 'PARTIAL_SUCCESS'
- 2. Search all preservation events with the eventOutcome 'FAILURE' and the eventType 'format identification'