Shanoir NG

Solr search specifications

Table of contents

[1 Context 1](#_Toc464640151)

[1.1 Project presentation 1](#_Toc464640152)

[1.2 Solr 1](#_Toc464640153)

[2 Indexing 2](#_Toc464640154)

[2.1 Data to index 2](#_Toc464640155)

[2.2 Indexing schedule 2](#_Toc464640156)

[3 Search page 2](#_Toc464640157)

[3.1 Search configuration 2](#_Toc464640159)

[3.1.1 Study name 3](#_Toc464640160)

[3.1.2 Subject name 3](#_Toc464640161)

[3.1.3 Dataset name 3](#_Toc464640162)

[3.1.4 Dataset creation date 3](#_Toc464640163)

[3.1.5 Examination comment 3](#_Toc464640164)

[3.2 Results 3](#_Toc464640165)

[3.2.1 Presentation 3](#_Toc464640166)

[3.3 Actions 4](#_Toc464640167)

[3.3.1 Dataset access 4](#_Toc464640168)

[3.3.2 Download DICOM data 4](#_Toc464640169)

# Context

## Project presentation

Following discussions on the way Shanoir Neurinfo is used by its users and the potential of Solr, this document describes how Shanoir Solr could:

* provide a surrogate to the Dataset and MyResearchStudies Shanoir menus with equivalent functionalities (in particular download of multiple series) and minimal development time : phase I
* provide improved functionalities with reasonable development time : phase II

## Solr

Solr (<https://lucene.apache.org/solr/>) is the popular, blazing-fast, open source enterprise search platform built on Apache Lucene™.

Solr is highly reliable, scalable and fault tolerant, providing distributed indexing, replication and load-balanced querying, automated failover and recovery, centralized configuration and more. Solr powers the search and navigation features of many of the world's largest internet sites.

Criteria are used to search data. There are two kinds of criteria:

* Text search
* Facet search

Solr also allows to sort results and to limit number of results.

Facet field values depend on search and displayed results.

# Indexing

## Data to index

Data to index are:

* Study identifier
* Study name
* Subject name
* Dataset identifier
* Dataset name
* Dataset type
* Dataset creation date
* Examination comment

## Indexing schedule

Data are indexed twice a day:

* 6 am
* 6 pm

# Search page

Search page is divided in two areas:

* Left area: search configuration
* Right area: results presentation and actions

A user can only access to datasets of studies the user has access to. This control is done on study identifier.

By default, no filter is applied and all datasets are displayed.

## Search configuration

Field facet values are limited to 5 by default. A “more” button allows to display all values. In this case, a “reduce” button limits values number to 5.

After facets selection, user clicks on “search” button to execute Solr request. While receiving response, results are displayed and facets are updated. Search buttons are displayed on the top and at the bottom of search configuration part.

Datasets could be filtered by these criteria:

* Study name
* Dataset creation date
* Subject name
* Dataset name
* Examination comment

### Study name

Study name criterion is a facet field. User is able to select zero, one or many study names. A checkbox is linked to each study name. Checkboxes are to the left of study name. They are vertically aligned.

If no study is selected, no filter on this criterion is applied.

### Subject name

Subject name criterion is a text field. User can combine keywords (AND, OR) to refine search.

A tooltip displays an information message explaining search and keywords.

Message: To refine search, “AND” and “OR” keywords could be used between terms.

### Dataset name

Dataset name criterion is a text field. User can combine keywords (AND, OR) to refine search.

A tooltip displays an information message explaining search and keywords.

Message: To refine search, “AND” and “OR” keywords could be used between terms.

### Dataset creation date

Dataset creation date criterion allows to search datasets between two dates. So it is composed of 2 search fields:

* Start date
* End date

These dates are defined with a calendar component.

### Examination comment

Examination comment criterion is a text field. User can combine keywords (AND, OR) to refine search.

A tooltip displays an information message explaining search and keywords.

Message: To refine search, “AND” and “OR” keywords could be used between terms.

## Results

### Presentation

Results are displayed in a tree. Structure of this tree is:

* Study name
  + Subject ID
    - Examination (date + comment)
      * Dataset name

Results data are sorted by default:

* Study name: alphabetical order
* Subject ID: alphabetical order
* Examination: ascending date
* Dataset: ascending creation date

User is able to warp and unwarp results.

A checkbox is to the right to each dataset name. User is able to select one or more datasets to execute actions.

Each study, subject and examination is linked to a checkbox (to the right of the item). Checking/unchecking one of these checkboxes will check/uncheck checkboxes of all corresponding datasets.

## Actions

User can access a dataset by clicking on its name on results tree.

After selecting one or many datasets, user can execute one of these actions:

* Download DICOM data

These actions are reachable in a drop-down list on the top of results part.

### Dataset access

Clicking on a dataset name redirects user to dataset view page in another window.

### Download DICOM data

If user choose many datasets, DICOM data are downloaded in a zip file.

DICOM zip files name format is: “[Subject ID]\_[Examination date]\_[Series description].zip”.

If user downloads datasets from one or many subjects, a unique zip file is downloaded. It contains one DICOM zip file by selected dataset.