Duplicate code detection for Cobol Tutorials

Technology preview-Version 1.0.1

Table of Contents

1 Introduction 2

2 Prerequisites 2

3 Setting the preferences 4

4 Duplicate code detection from a Cobol file 6

4.1 Search duplicates in the current file 6

4.2 Search duplicates from a selection 6

5 Duplicate code detection in multiple files 7

5.1 Create a duplicate code project from Eclipse 7

5.1.1 From the Resource perspective, from the contextual menu of the Project Explorer 7

5.1.2 From the Duplicate Code Project view 8

5.2 Build a project from the Duplicate Code Projects view 11

5.3 Duplicate Code Projects list from a server 12

5.4 Search requests 12

5.4.1 Where to ask for duplicates? 12

5.4.2 The search dialog 14

6 View the duplicates 17

6.1 The duplicates view 17

6.2 The fragments view 19

6.3 The duplicate details view 21

7 Compute the refactoring options 23

1 Introduction

This documentation includes a tutorial to learn the steps to use the Duplicate Code Detection tool in IBM Developer for z/OS.

Learning objectives:

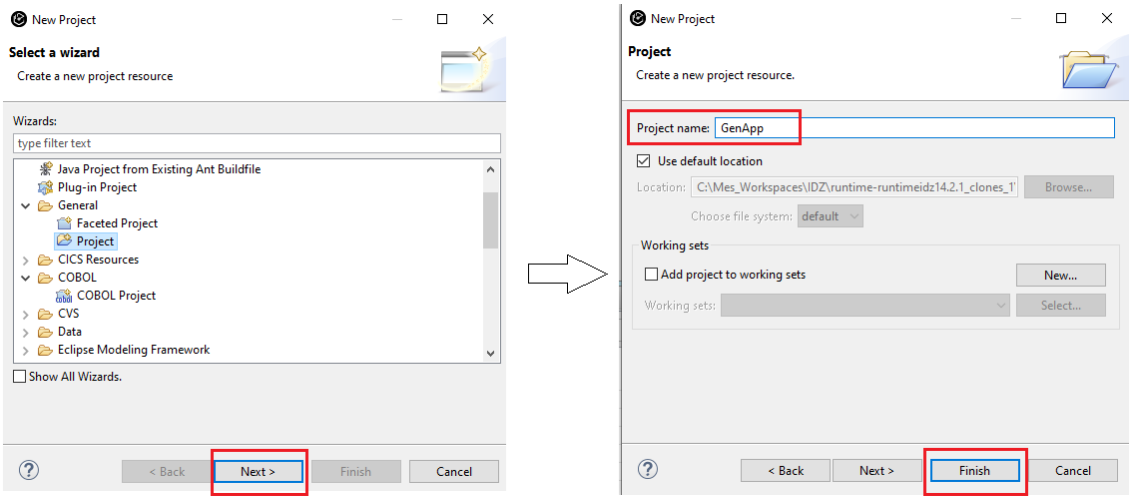
- Setting the preferences
- Duplicate code detection in a single file
- Duplicate code detection in multiple files: create a project, build it and search for the duplicates
- View the duplicates
- View the refactoring options

2 Prerequisites

CLIENT

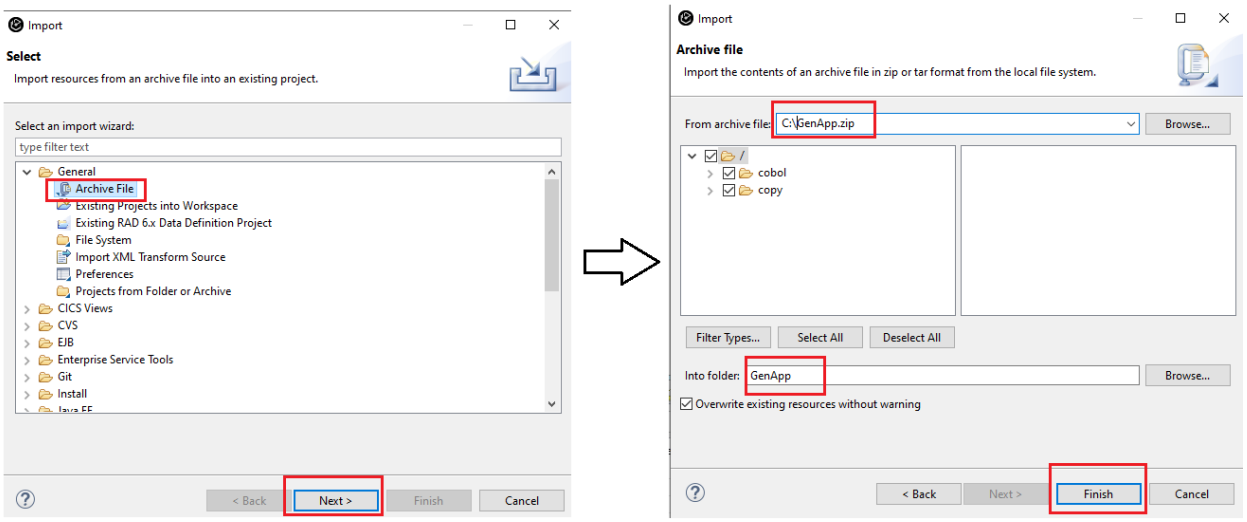
This tutorial is based on the **GenApp** application.
To load your workspace with this application:

- Download GenApp.zip.
- From the Resource perspective of IDz, from the contextual menu of the Project Explorer:
    - Select **New > Project**, the New Project wizard opens.

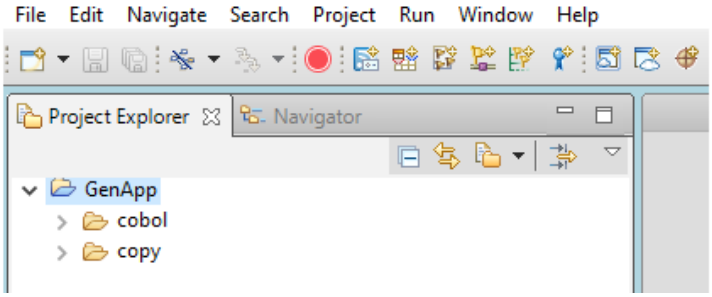


- Import the GenApp content:

  From the GenApp project, right-click and select **Import**, the import wizard opens.



- The **GenApp** project is now created in the workspace:



SERVER

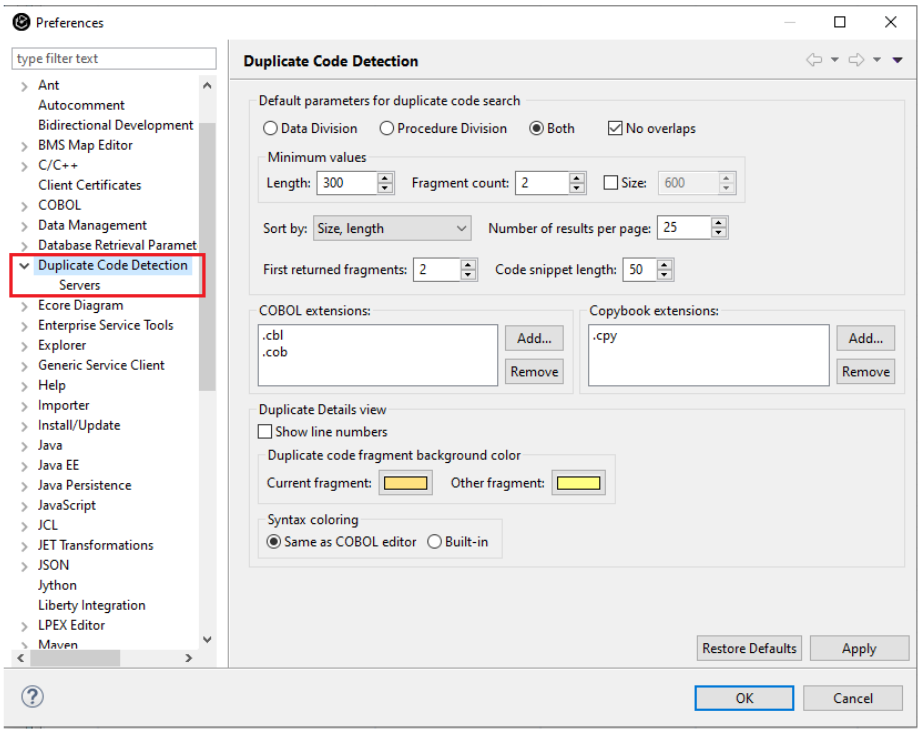
A Liberty server is required if you want to use the feature with Git or EWM. Otherwise, no server is required.

Start the server that contains the server part of the Duplicate Code Detection feature and define it in the preferences (refer to DCD Installation guide).

During the first access to the server you will have to enter your credentials.

3 Setting the preferences

You can set preferences that control the appearance and operation of Duplicate Code Detection.
Open the **Preferences** window then **Duplicate Code Detection**.



Default parameters for duplicate code search

Data Division, Procedure Division or both

Indicates the scope of the search. Both by default.

No overlaps

If selected, the duplicates search finds fewer results by making only 1 duplicate among a group of overlapping duplicates.

Number of results per page/minimum values

- Length**:** 300 by default.
- Fragment count**:** 2 by default.

Sort by

Indicates what duplicates must be returned in priority.

- Length: number of significant characters of a duplicate.
- Fragments count: number of repetitions of a duplicate.
- Size, length: product of the two criteria. Default.
- Size, fragments count: product of the two criteria.

Number of results per page

25 by default.

First returned fragments

Number of first returned fragments in the fragments view. 2 by default.

Code snippet length

Length of the snippet in the results view. 50 characters by default.

COBOL extensions

Specifies the COBOL extensions used for the analysis of COBOL Programs.
You can add or delete via the **Add...** and **Remove** buttons.

Copybook extensions

Specifies the Copybook extensions used for the analysis of copybooks.
You can add or delete via the **Add...** and **Remove** buttons.

Duplicate detail view presentation

Show line numbers

Check this box to see lines numbers in the detail view.

Duplicate code background color

Current fragment

Personalize the colors used to show the current duplicates.

Other fragment

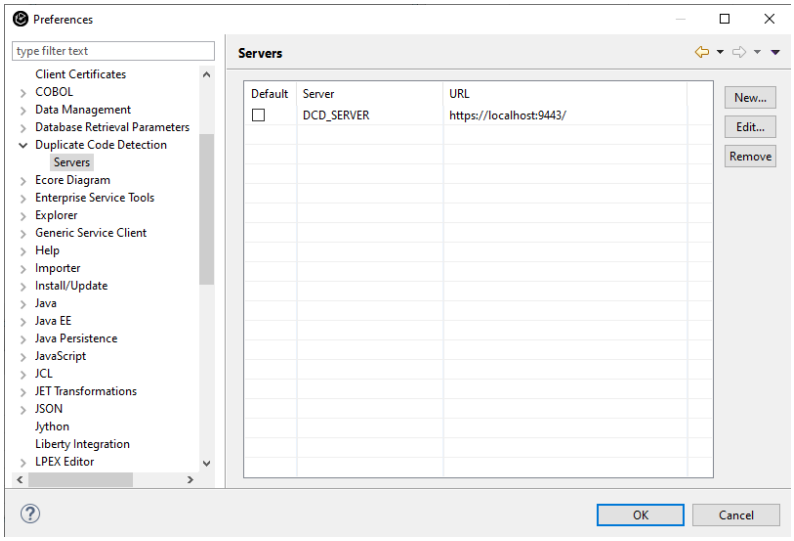
Personalize the colors used to show the duplicates.

Syntax coloring

The syntax coloring may be the same that the Cobol editor (default) or managed by the DCD feature.

Server preferences

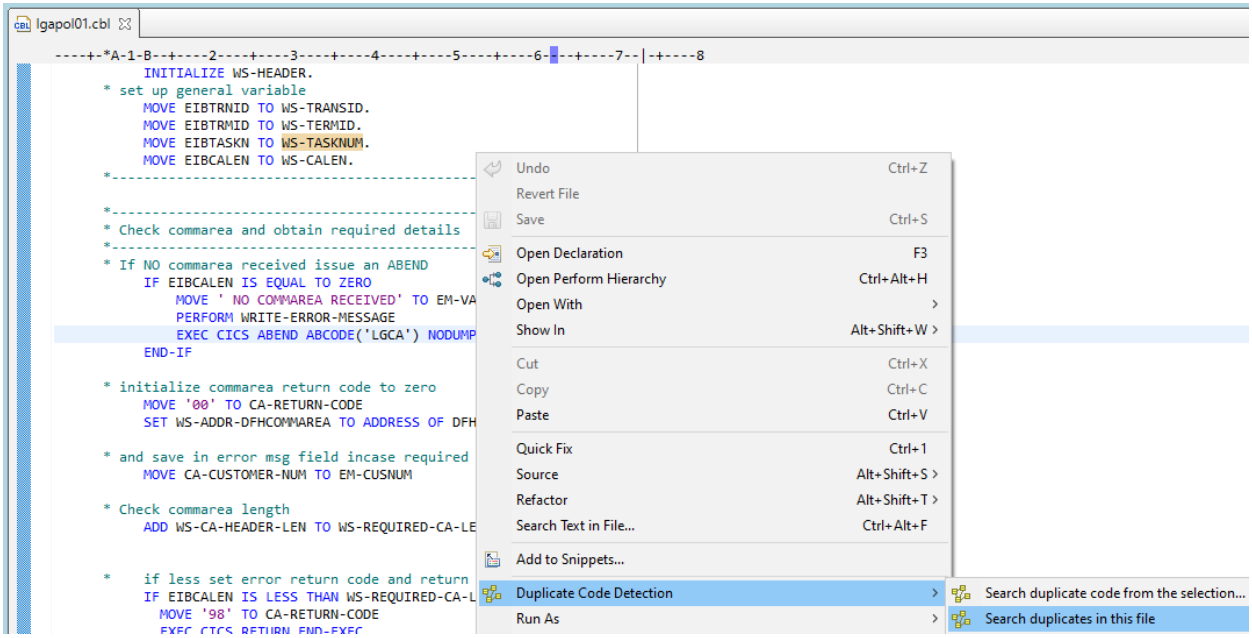
Refer to DCD Installation guide section Configure the client.



4 Duplicate code detection from a Cobol file

4.1 Search duplicates in the current file

This action is completely local. No server is required.
Open a COBOL program or a COBOL copybook with the COBOL Editor.
Right-click and choose **Duplicate Code Detection** menu and **Search duplicates in this file**.



The **Duplicates** view takes the focus. Refer to the section View the duplicates

4.2 Search duplicates from a selection

Open a COBOL program or a COBOL copybook with the COBOL Editor.
Select a piece of code and right-click and choose **Duplicate Code Detection** menu and **Search duplicate code from the selection**.
Refer to the section Search from the Cobol Editor

5 Duplicate code detection in multiple files

5.1 Create a duplicate code project from Eclipse

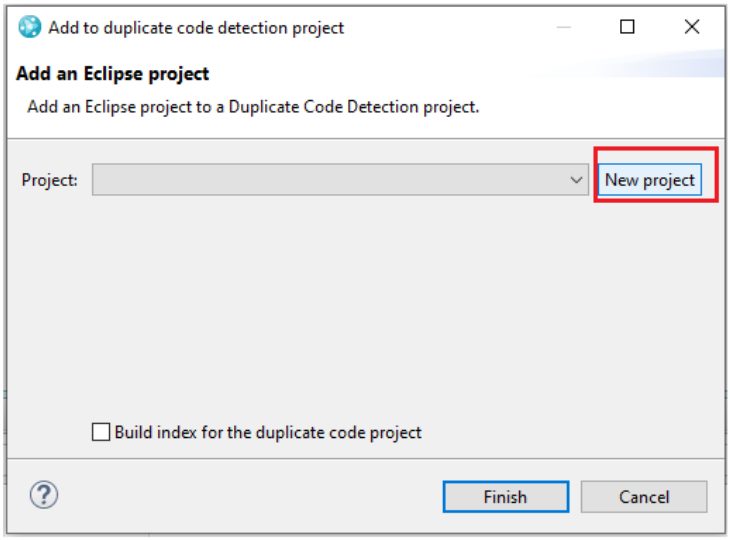
A Duplicate Code Detection project allows to define a container for a set of Cobol source artifacts. This container is fully copied (will all Cobol artifacts) in the Duplicate Code Detection Storage and is used as a reference for all requests on multiple files.

This action is only available with a local workspace. It is not available with a server.

There are two ways to create a duplicate code project.

5.1.1 From the Resource perspective, from the contextual menu of the Project Explorer

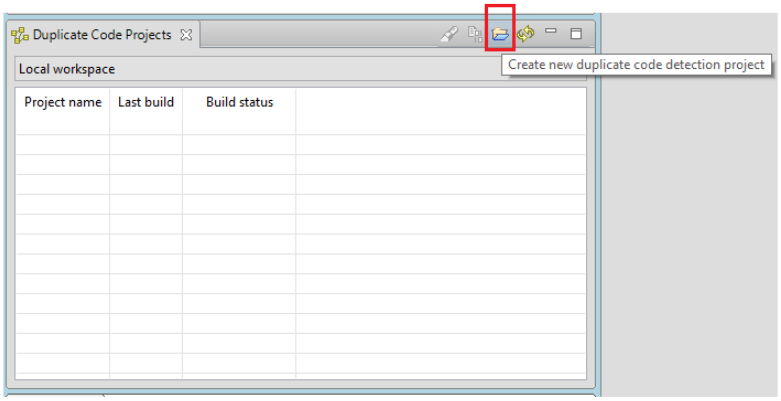
To associate a COBOL project to a new duplicate code project, right-click on the **GenApp** project and select **Duplicate Code Detection > Add Eclipse Project…**
The **Add in Duplicate code detection project** wizard opens.



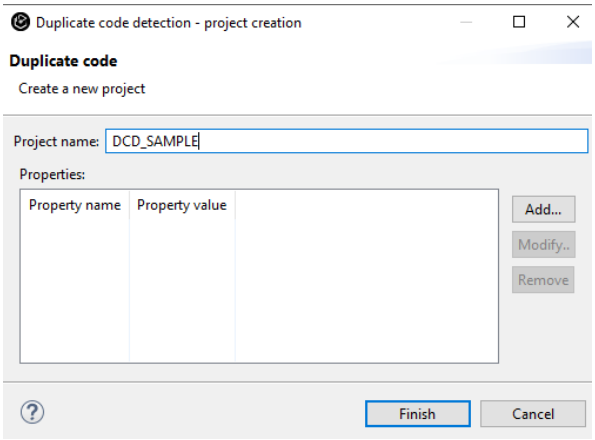
You can choose to build the project in the same request by checking the corresponding box.
Click on the **New project** button to set up the project name. (go to the section Set up the project name and the properties).

5.1.2 From the Duplicate Code Project view

Select **Window > Show view > Duplicate Code Detection > Duplicate Code Project** view.
Click the **Create new duplicate code detection project** button.



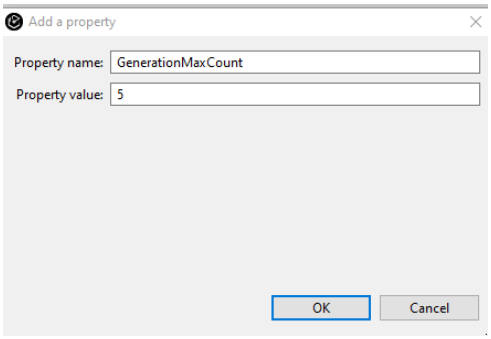
5.1.2.1 Set up the project name and the properties



Give a name to the project.

When you create a project, you can add some properties with the **Add...** button.
These properties cannot be updated in this version of the feature.

By default, the number of kept builds is 3. If you want to increase this number, you can add this property: `GenerationMaxCount`



Press **OK** then **Finish**.

Properties of a project are visible in the view "Properties"

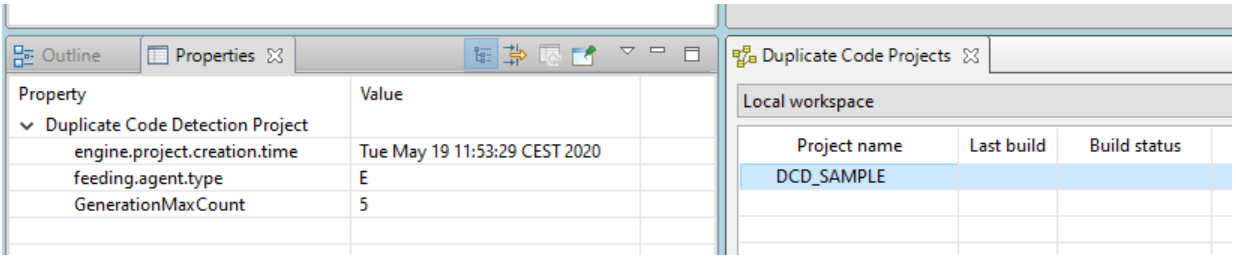
The properties are technical data, but they can be viewed:

feeding.agent.type

E	Eclipse
G	Git
R	EWM (RTC)

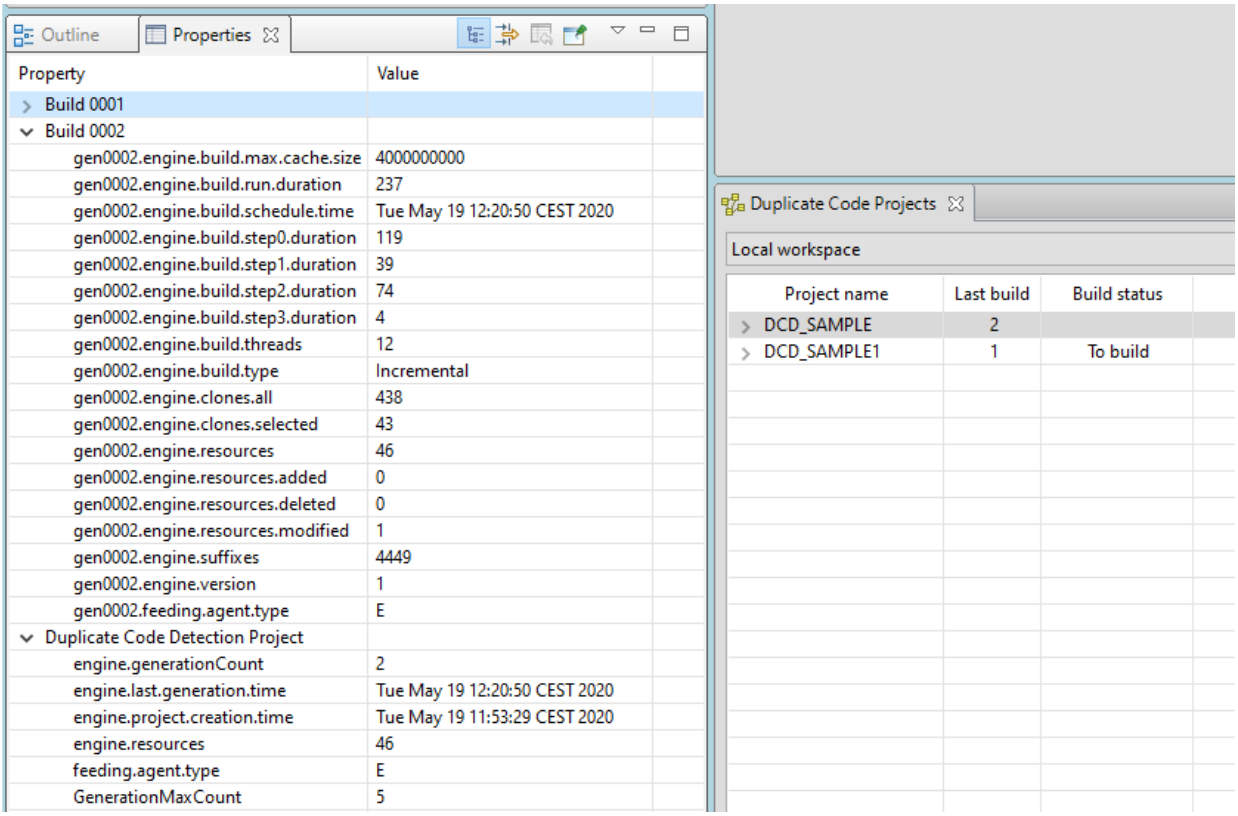
Examples:

1. Creation of a project with modification of the number of builds.



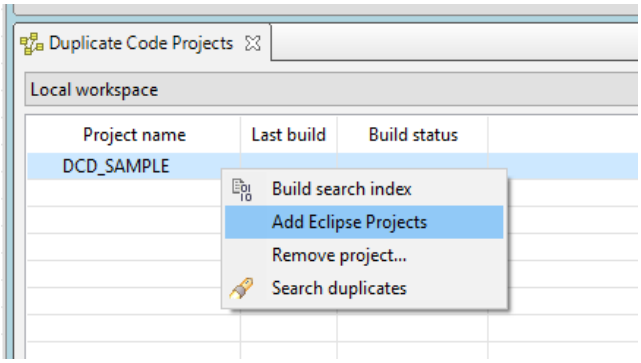
2. Project with several builds

Properties on build give information on duration, number of added/modified/deleted resources…

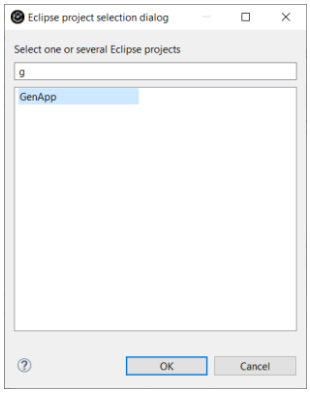


5.1.2.2 Associate an Eclipse Project

In the case of a creation from the **Duplicate Code Project** view, you must associate this project with one or several Eclipse projects that contains COBOL sources. So, right-click the line of the duplicate code project and select **Add Eclipse Projects.**

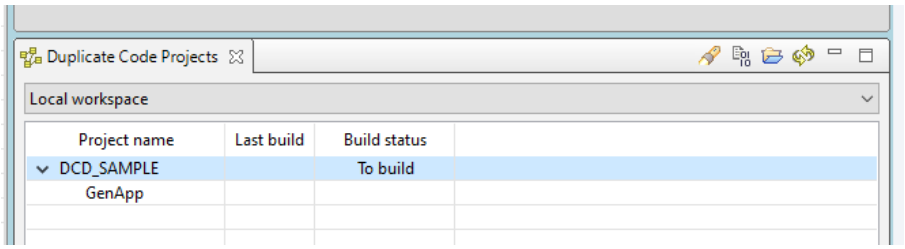


In **Eclipse project selection** dialog, you can either enter the name of an existing Eclipse project, or you can enter * to display all the existing Eclipse projects and to select one of them.

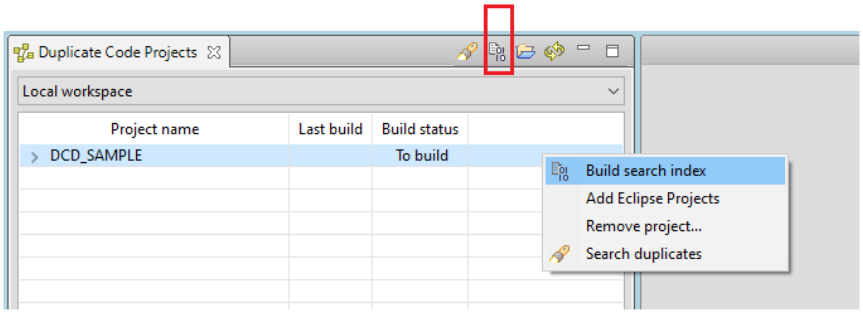


5.2 Build a project from the Duplicate Code Projects view

This view shows all the duplicate code projects. The status is marked as “To build” if the project must be built to update the index files on the duplicate code detection data base.
The last build rank is incremented upon each build.



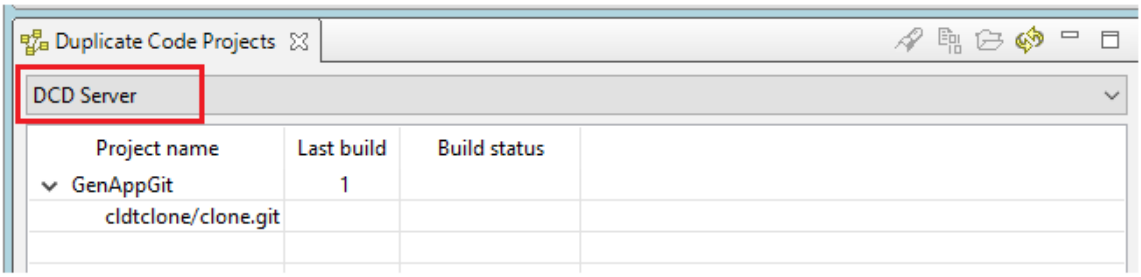
When a project is desynchronized and therefore marked “To build”, you can build it by right-clicking a project and selecting **Build search index** or click on **Build search index** button in the toolbar.



The duplicate code project is stored in the metadata of the workspace.

5.3 Duplicate Code Projects list from a server

If you have defined servers URL in the preferences, the combo list is populated with the “Local workspace” and server names.
You can select a server to see the projects in this server.



For the projects created and built from the SCM script, no action is available in the view except search duplicates.

Now, you can start the search requests.

5.4 Search requests

5.4.1 Where to ask for duplicates?

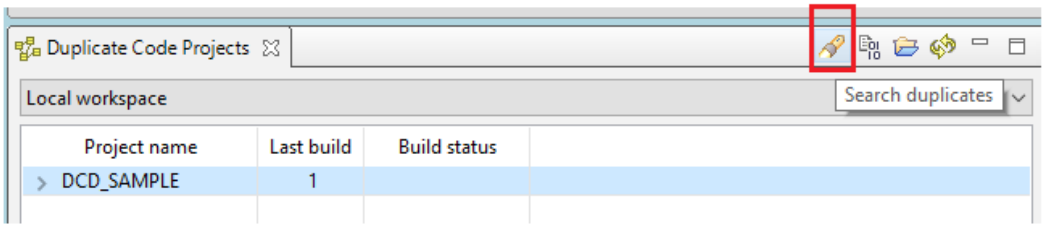
Duplicates can be searched for from multiple locations.

5.4.1.1 Search from the Workbench

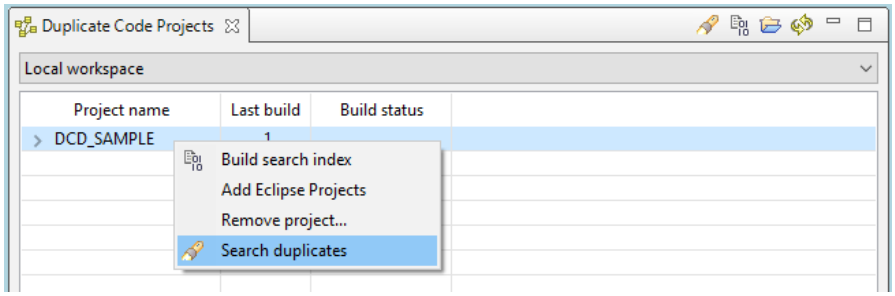
- In the Workbench toolbar, click the **Search** button.
- Go to **Duplicate Code Search** tab.

5.4.1.2 Search from the Duplicate Code Projects view

- In the toolbar, click the **Search duplicates** button



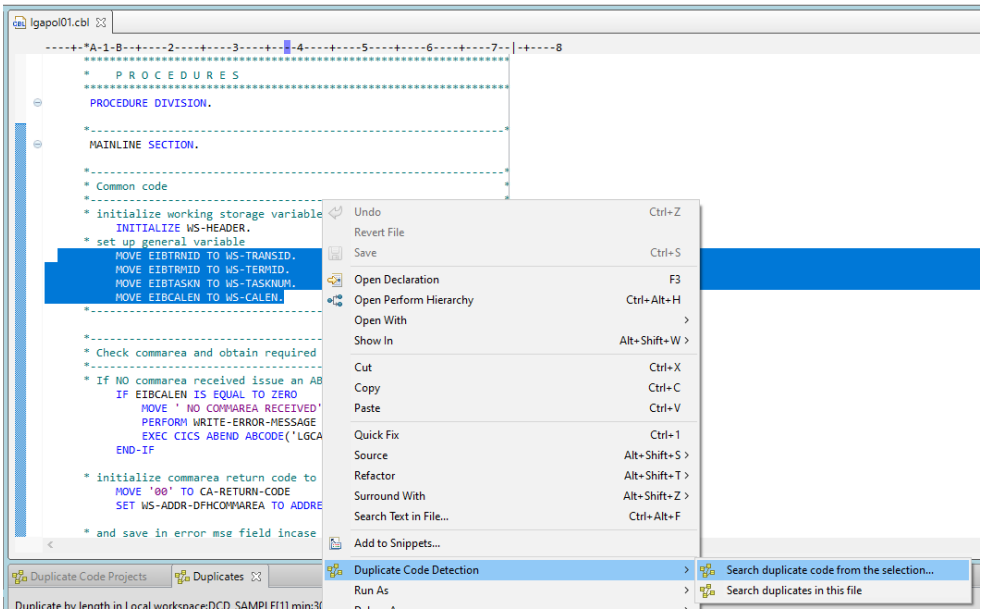
- In the contextual menu, choose **Search duplicates**.



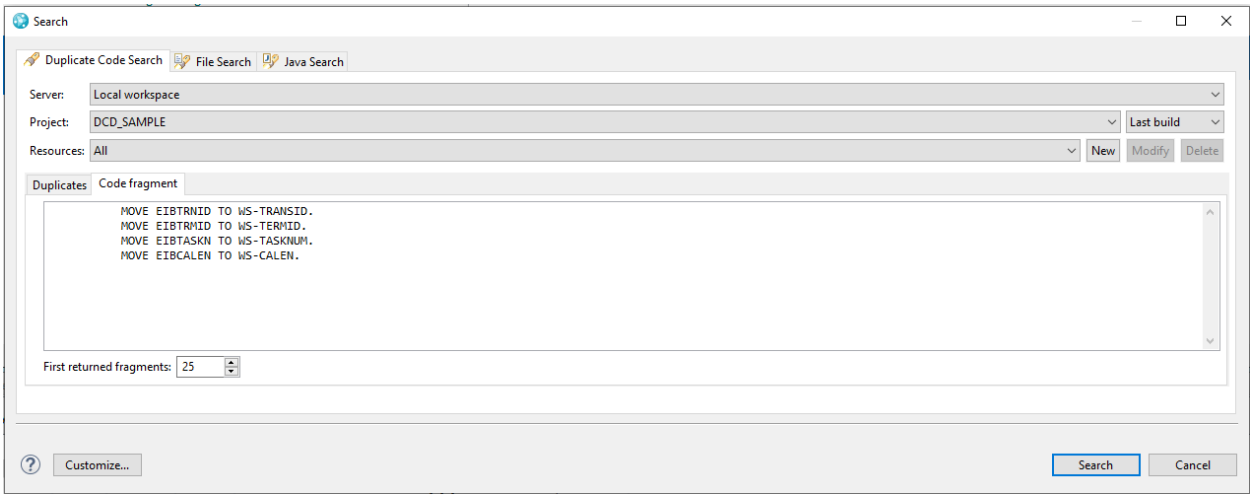
The **Search** dialog takes focus and the fields are initialized with the current data of the item.

5.4.1.3 Search from the Cobol Editor

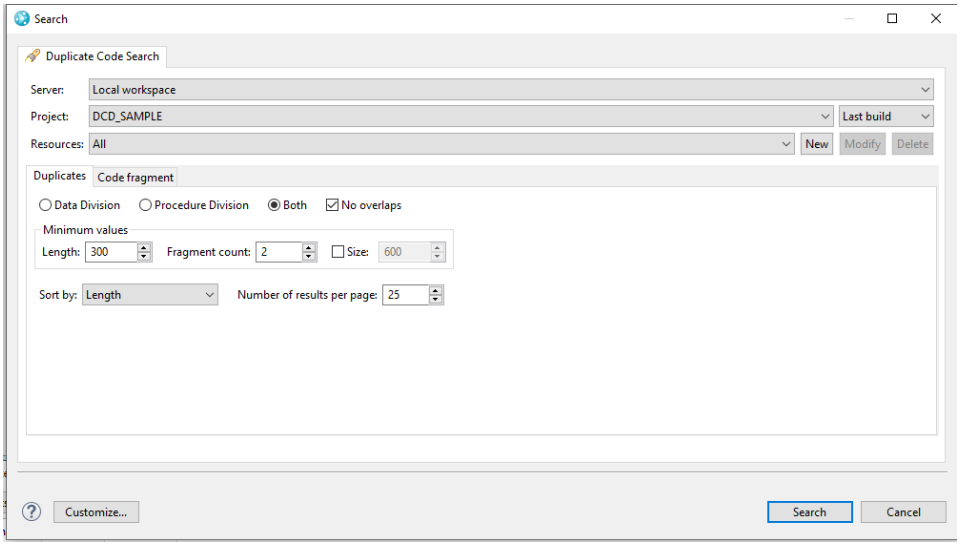
Select a piece of code and right-click and choose **Duplicate Code Detection** menu and **Search duplicate code from the selection**.



The **Search** dialog takes focus and the fields are initialized with the selection.



5.4.2 The search dialog



You can search in

Server

Select “Local workspace” because your duplicate code project is stored in your local workspace.

If the duplicate code project was created from an SCM feeding agent, you could select the liberty server where the duplicate code project was stored.

You must specify the project, the build rank you want to search in and if you want to search in the Data Division, the Procedure Division or both.

Results presentation

The values are initialized with the default values defined in the preferences.

Sort by

Indicates what duplicates must be returned in priority.

- Length: number of significant characters of a duplicate.
- Fragments count: number of repetitions of a duplicate.
- Size, length: product of the two criteria.
- Size, fragments count: product of the two criteria.

Number of results per page

Minimum length

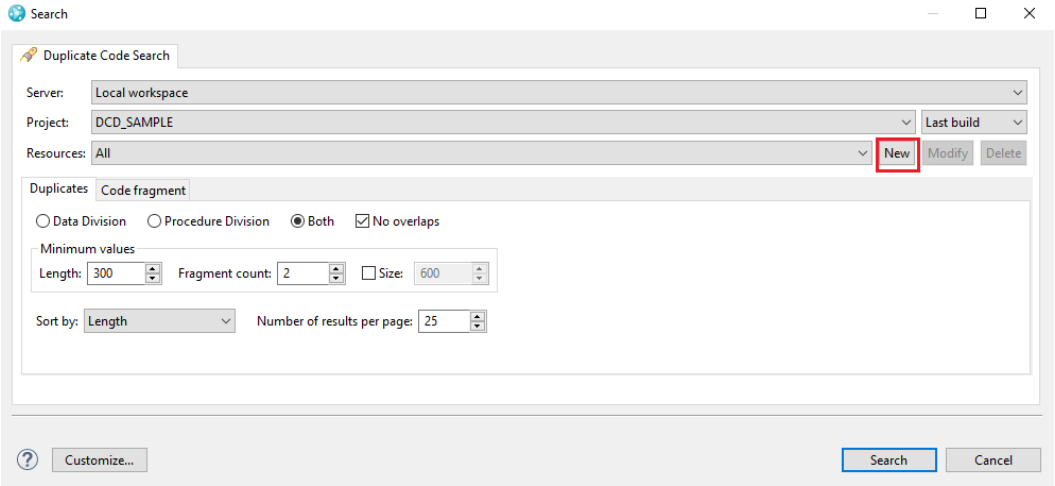
Recommended to put a minimum length to have significant duplicates.

Minimum fragment count

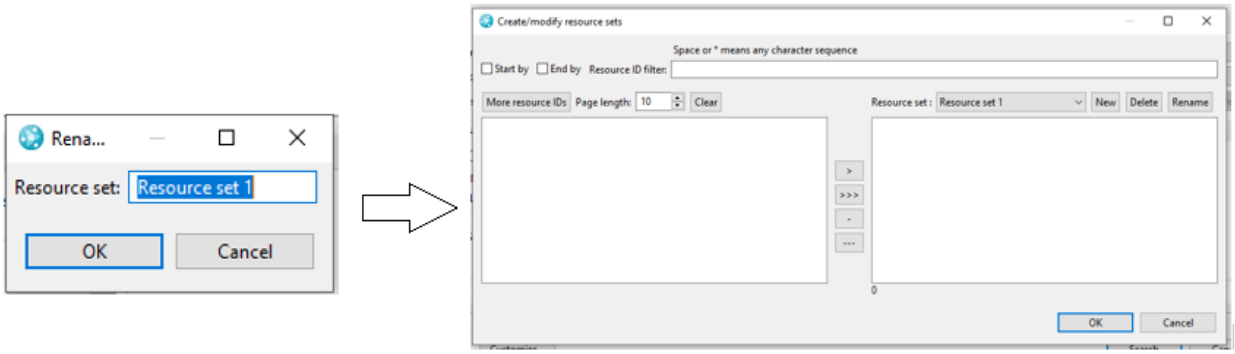
You can search for:

- *All duplicates*
- *All duplicates in a resource set*

Click on the **New** button



Choose a name for the resource set, then **OK**.



Start by, End by, Resource ID filter

Used to search resources ID.

Enter the character string that you want to search for. This string represents the whole or partial name of the element that the search is run on.

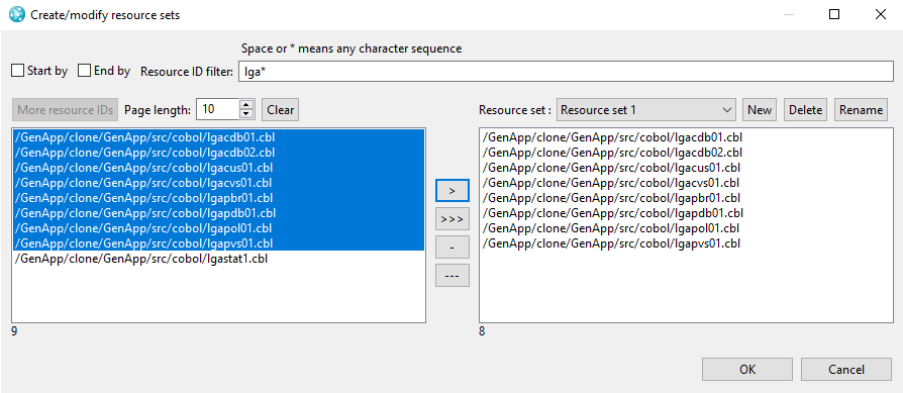
Then click on **More resource Ids** to start the search.

For examples,

foo in the resource ID allows to list all the resources ID that contains **foo**.

foo BAR in the resource ID allows to list all the resources ID that contains **foo** then **BAR**.

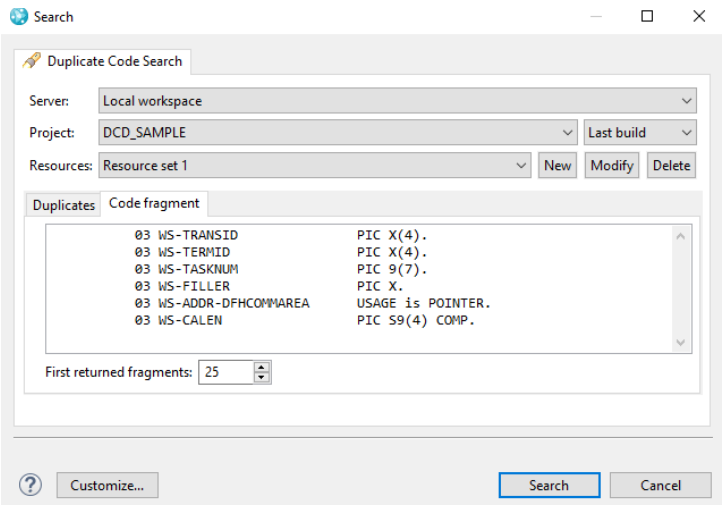
The results are displayed in the left and you choose what are the resource IDs to add to the resource set: all (**>>>**) or only the selected ones (**>**).



Then click on **OK** button to return to the **search** dialog.

- *Duplicates of code fragment*

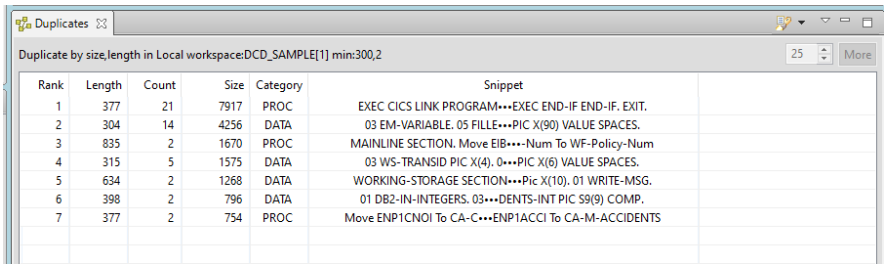
A piece of Cobol code must be added. Be careful to keep the left margin. Only the duplicates that correspond to the code are returned.
Search on fragments are also available with a resource set search.



Press **Search** to start the process. The **Duplicates** view takes the focus.

6 View the duplicates

6.1 The duplicates view



Columns

Rank

The duplicates are numbered in the view.

Length

Number of significant characters. Comments, margins and spaces are not counted.

Count

Number of times a code fragment is repeated.

Size

Length * Count.

Category

PROC Procedure Division.
DATA Data Division.

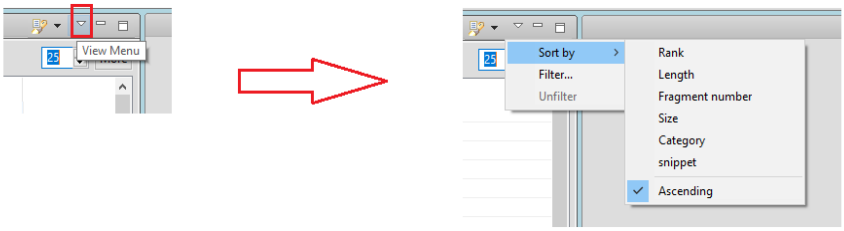
Snippet

A short description of the fragment. The length is specified in the preferences.

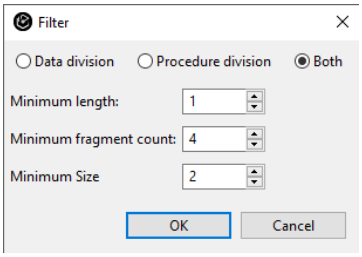
The line above the table shows the parameters of the request. This line appears in the history of the requests that can be displayed with the button.

The page displays the number of duplicates specified in the request. To see more results, use the **More** button.

You can sort the results differently by clicking the column headers or by selecting a sort criterion in the **Sort** menu of the view toolbar.

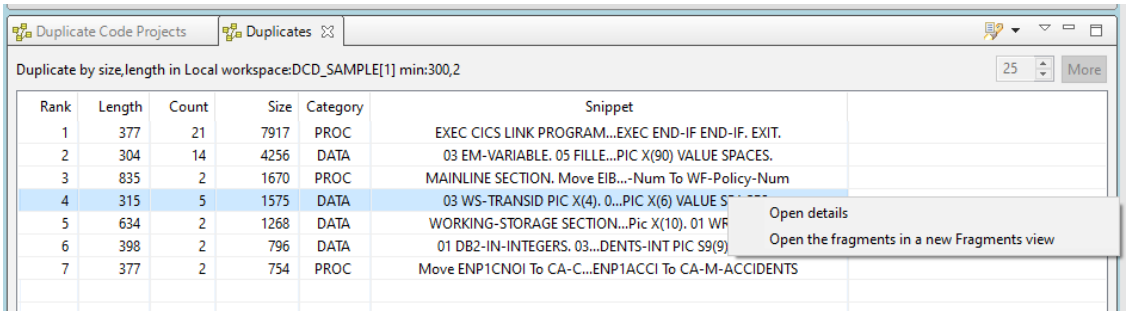


You can filter the results in the **Filter** dialog.
For example, filter on Procedure division.

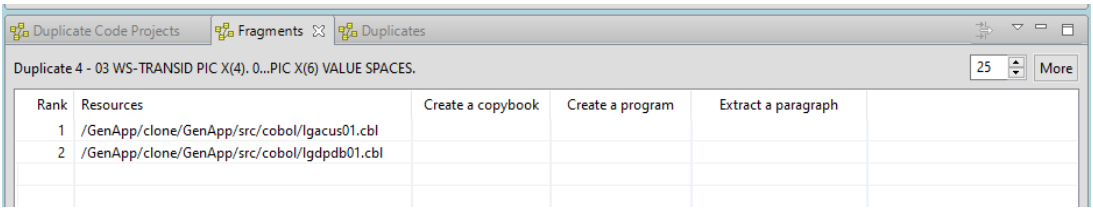


When a filter is applied, the number of the duplicates displayed relatively to the total number of duplicates is indicated above the table.

From each line, you can **Open the Duplicate Details** view or **Open the fragments in a new Fragments view.**



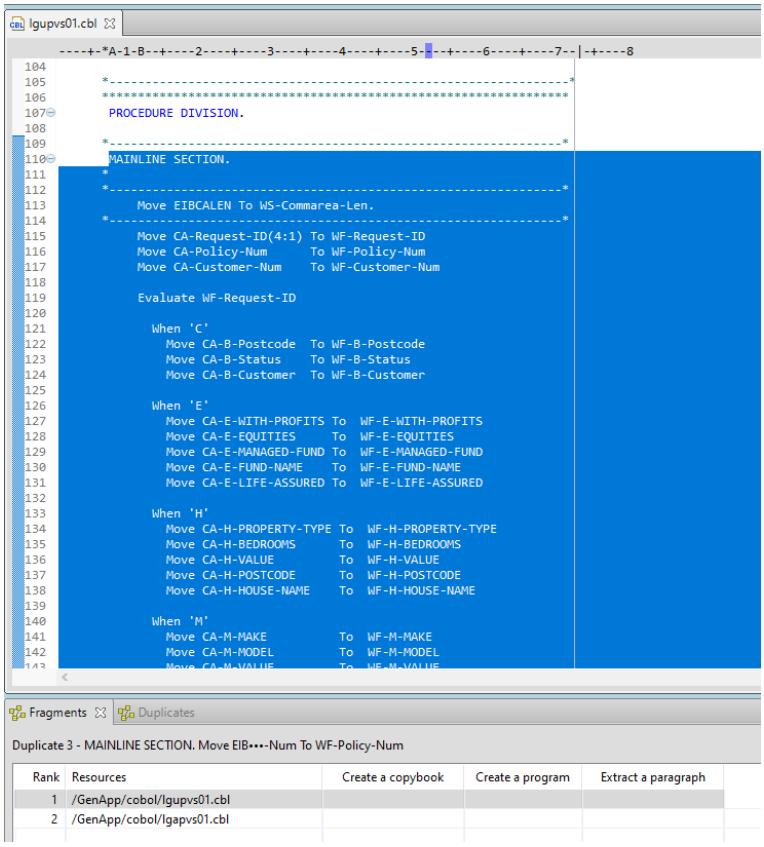
6.2 The fragments view



The page displays the number of first fragments specified in the preferences. To see more results, use the **More** button.

If a fragments view is opened, its content is dynamically update when the user selects another duplicate in the duplicates view. The link is activated for only the first instance of the fragments view.

If you double-click a file, you open it in the standard **COBOL** editor. The duplicate fragment is highlighted.



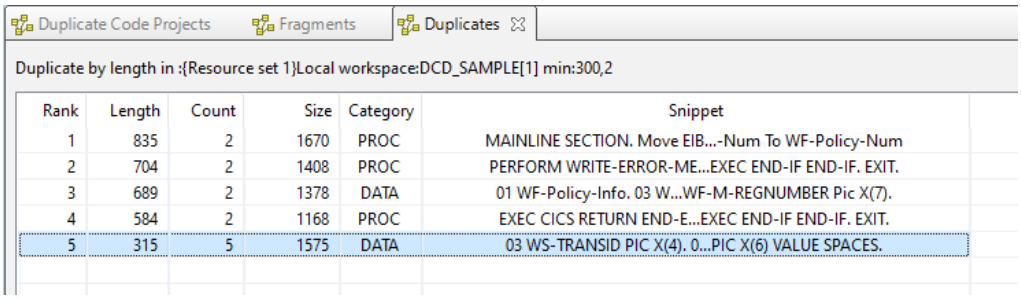
If you right-click a duplicate and select **Open Details,** the **Duplicate details** view opens.

Resource set filtering

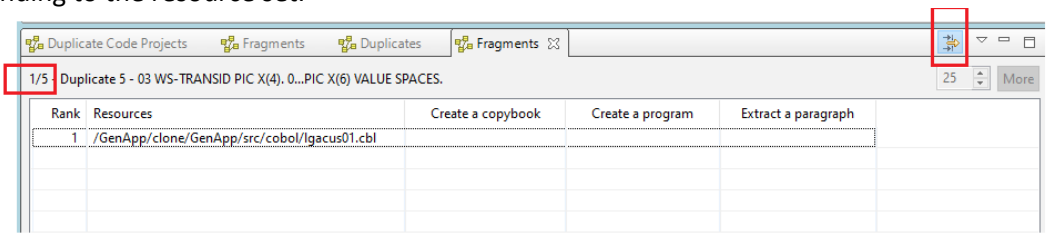
When the search is done with a resource set, the fragments count displayed in the **duplicates** view is the total fragments count, not the filtered fragments count.

Example:

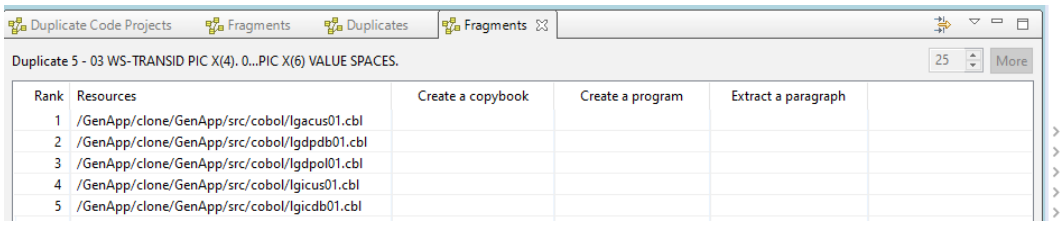
The following duplicates view is the result of a search with a resource set.

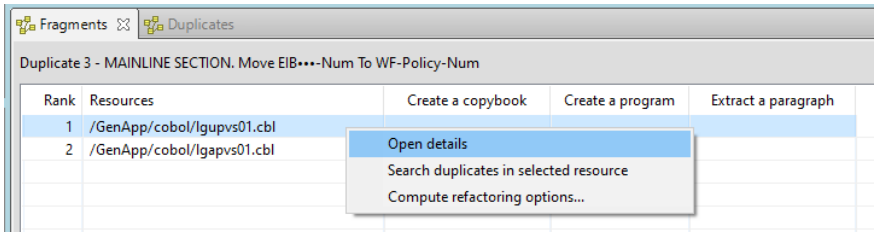


There are 5 fragments for the duplicates number 5. But the fragments view shows only one fragment corresponding to the resource set.



The button is now available. By clicking on it, the filter can be disabled or enabled.

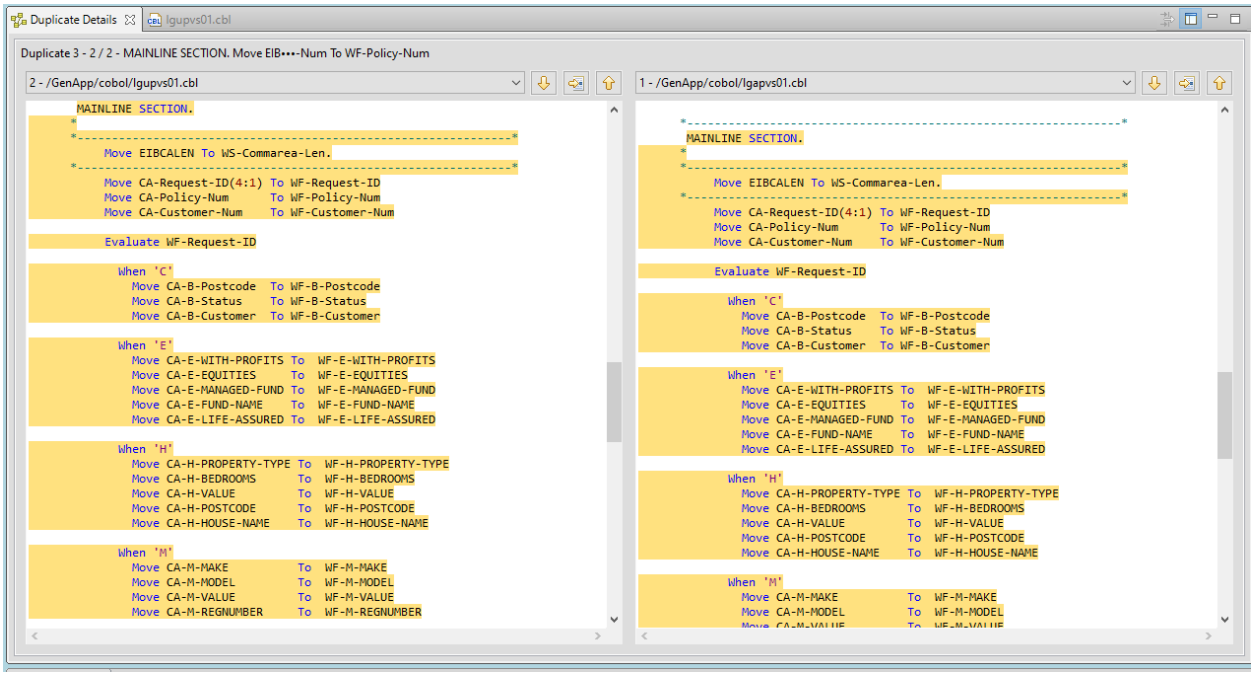




6.3 The duplicate details view

This view displays two COBOL editors side by side. Each one highlights one of the duplicates uses of the fragment. In the case below, since the two uses correspond to 2 files, each COBOL editor displays a fragment in one file. If the duplicate is found in the same file, then the same file is displayed in both editors.

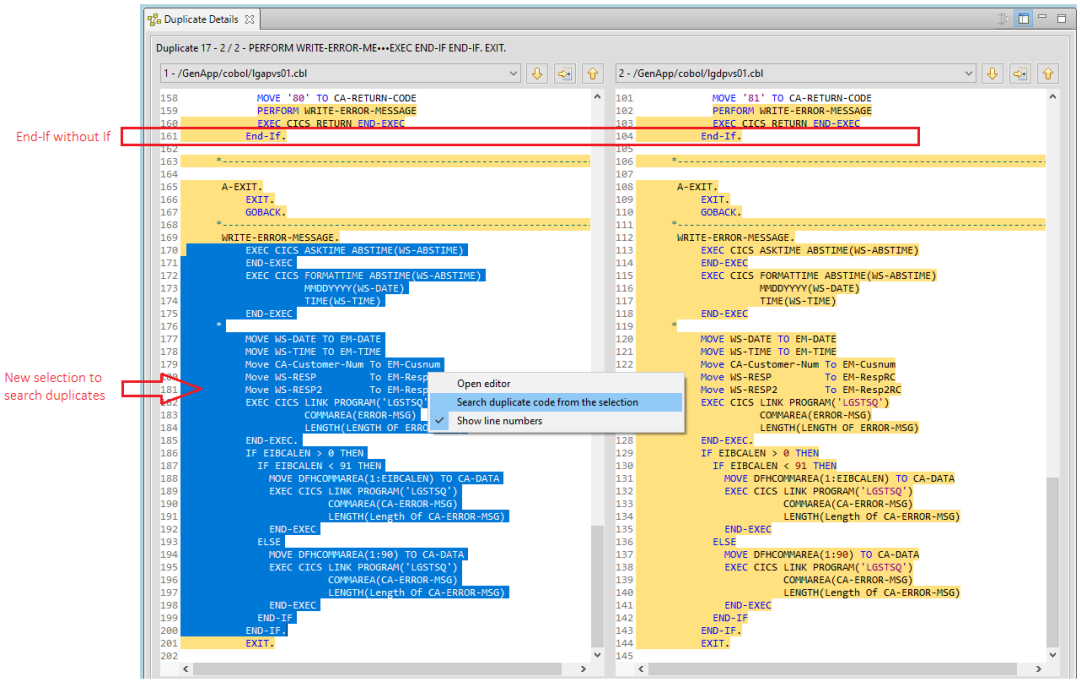
If a duplicate is found in more than 2 files, you can change the files that are displayed.



The color used to highlight the code can be updated in the preferences. Refer to section Setting the preferences Setting the preferences

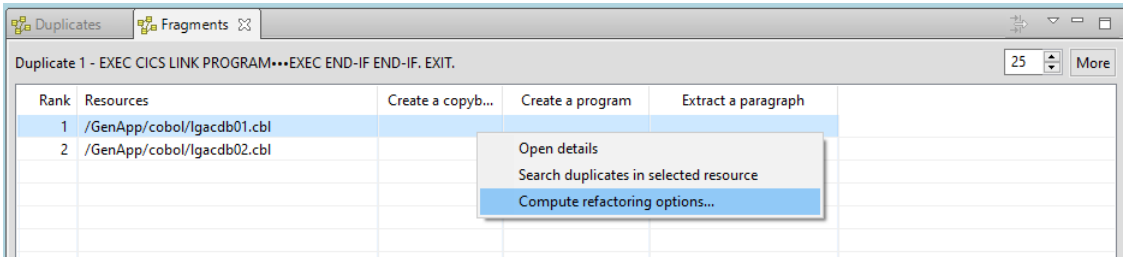
You can see lines numbers by right-clicking in an editor then **Show line numbers**.

A duplicate corresponds to an entire statement. For example, a PERFORM alone is not considered as a duplicate.
A duplicate is the maximum duplicate code found. It cannot be expanded more.
This is the reason why duplicates might not be syntactically balanced.
In the following example, the duplicate (in orange) includes **End-If** but does not include the **If** statement.
If you want to narrow down to a balanced fragment, you can select a syntactically balanced group of code lines (in blue), right-click and **Search duplicate code from the selection**.

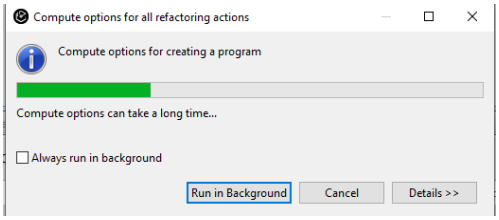

End-If without If
New selection to search duplicates
Duplicate Details
Duplicate 17 - 2 / 2 - PERFORM WRITE-ERROR-ME•••EXEC END-IF END-IF. EXIT.
1 - /GenApp/cobol/lgapvs01.cbl
2 - /GenApp/cobol/lgdpvs01.cbl
Open editor
Search duplicate code from the selection
Show line numbers

7 Compute the refactoring options

Now, you can request which solution is the best to remove a duplicate. To do that, on the **Fragments** view, right-click on a duplicate or a file and select **Compute refactoring options**. The selection of several duplicates is also possible.

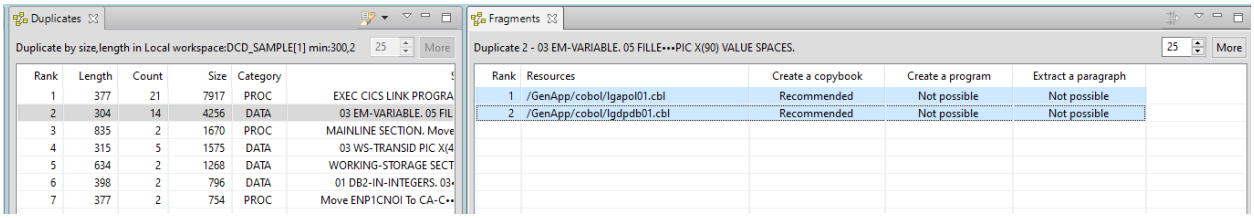


The processing takes over the IDz refactoring controls as they are applied in the COBOL editor (refer to the IDz documentation on the refactoring Idz Refactor).
The process can take a long time depending on the number of files to process. A dialog opens and you have the possibility to run it in the background. During the process, it is not possible to refactor the code in a program via the Cobol editor.



When it is finished, the status is displayed:

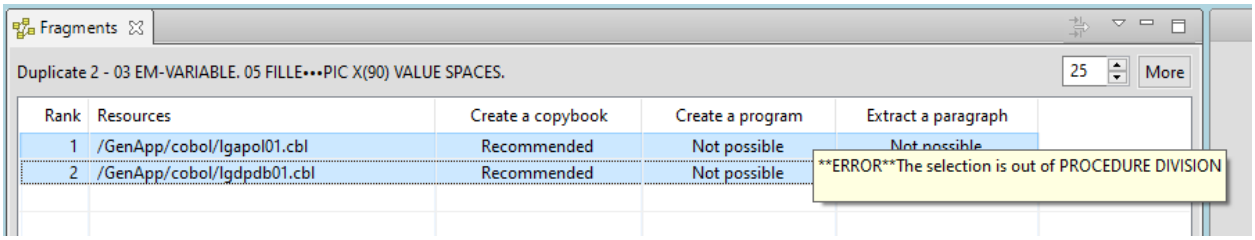
- Recommended
- Possible
- Not possible: an error blocks the refactoring.
- Desynchronized: the files in the duplicate code detection server and in the local workspace are desynchronized. The project requires regeneration.
- Not applicable: refactoring is not applicable on a copybook.



h

You can see that for the duplicates above, **Create a copybook** is recommended because the duplicates exist in DATA DIVISION.

If a refactoring option is not possible, a tooltip displays the reasons.



Refer to the video for an example of refactoring.