

# How To... Enable Automatic Duplicate Check for Resources in KM Repositories

Applicable Releases:

SAP NetWeaver 7.0

IT Practice:

User Productivity Enablement

IT Scenario:

Enterprise Knowledge Management

Version 1.0

December 2008

© Copyright 2008 SAP AG. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors.

Microsoft, Windows, Outlook, and PowerPoint are registered trademarks of Microsoft Corporation.

IBM, DB2, DB2 Universal Database, OS/2, Parallel Sysplex, MVS/ESA, AIX, S/390, AS/400, OS/390, OS/400, iSeries, pSeries, xSeries, zSeries, z/OS, AFP, Intelligent Miner, WebSphere, Netfinity, Tivoli, Informix, i5/OS, POWER, POWER5, OpenPower and PowerPC are trademarks or registered trademarks of IBM Corporation.

Adobe, the Adobe logo, Acrobat, PostScript, and Reader are either trademarks or registered trademarks of Adobe Systems Incorporated in the United States and/or other countries.

Oracle is a registered trademark of Oracle Corporation.

UNIX, X/Open, OSF/1, and Motif are registered trademarks of the Open Group.

Citrix, ICA, Program Neighborhood, MetaFrame, WinFrame, VideoFrame, and MultiWin are trademarks or registered trademarks of Citrix Systems, Inc.

HTML, XML, XHTML and W3C are trademarks or registered trademarks of W3C®, World Wide Web Consortium, Massachusetts Institute of Technology.

Java is a registered trademark of Sun Microsystems, Inc.

JavaScript is a registered trademark of Sun Microsystems, Inc., used under license for technology invented and implemented by Netscape.

MaxDB is a trademark of MySQL AB, Sweden.

SAP, R/3, mySAP, mySAP.com, xApps, xApp, SAP NetWeaver, and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries all over the world. All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

These materials are subject to change without notice.

These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

These materials are provided "as is" without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

SAP shall not be liable for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials.

SAP does not warrant the accuracy or completeness of the information, text, graphics, links or other items contained within these materials. SAP has no control over the information that you may access through the use of hot links contained in these materials and does not endorse your use of third party web pages nor provide any warranty whatsoever relating to third party web pages.

SAP NetWeaver "How-to" Guides are intended to simplify the product implementation. While specific product features and procedures typically are explained in a practical business context, it is not implied that those features and procedures are the only approach in solving a specific business problem using SAP NetWeaver. Should you wish to receive additional information, clarification or support, please refer to SAP Consulting.

Any software coding and/or code lines / strings ("Code") included in this documentation are only examples and are not intended to be used in a productive system environment. The Code is only intended better explain and visualize the syntax and phrasing rules of certain coding. SAP does not warrant the correctness and completeness of the Code given herein, and SAP shall not be liable for errors or damages caused by the usage of the Code, except if such damages were caused by SAP intentionally or grossly negligent.

#### Disclaimer

Some components of this product are based on Java™. Any code change in these components may cause unpredictable and severe malfunctions and is therefore expressly prohibited, as is any decompilation of these components.

Any Java™ Source Code delivered with this product is only to be used by SAP's Support Services and may not be modified or altered in any way.

## Document History

<b>Document Version</b>	<b>Description</b>
-------------------------	--------------------

---

1.00	First official release of this guide
------	--------------------------------------

---

## Typographic Conventions

Type Style	Description
<i>Example Text</i>	Words or characters quoted from the screen. These include field names, screen titles, pushbuttons labels, menu names, menu paths, and menu options.  Cross-references to other documentation
<b>Example text</b>	Emphasized words or phrases in body text, graphic titles, and table titles
Example text	File and directory names and their paths, messages, names of variables and parameters, source text, and names of installation, upgrade and database tools.
<b>Example text</b>	User entry texts. These are words or characters that you enter in the system exactly as they appear in the documentation.
< <b>Example text</b> >	Variable user entry. Angle brackets indicate that you replace these words and characters with appropriate entries to make entries in the system.
EXAMPLE TEXT	Keys on the keyboard, for example, F2 or ENTER.

## Icons





Icon	Description
	Caution
	Note or Important
	Example
	Recommendation or Tip

Table of Contents

- 1. **Scenario**..... 1
- 2. **Background Information**..... 1
- 3. **Prerequisites** ..... 1
- 4. **Step-by-Step Procedure**..... 3
  - 4.1 Obtain and Deploy KM Report..... 3
  - 4.2 Prepare Duplicate Check (*Initialize* Command) ..... 3
  - 4.3 Duplicate Check (*Move Duplicates* Command)..... 6
- 5. **Appendix** ..... 8

## 1. Scenario

You want to keep your repository free of duplicate documents? Then you are in the right place. This How-To-Guide introduces a solution based on KM Reporting for this concern.

The KM Report detects existing duplicates in a repository and either just lists them or additionally moves them to a prior specified location. Once identified one may decide how to proceed with the duplicate documents.

### Important

The KM Report lists respectively moves all duplicate documents except the one with the **newest** *last modified* date.

### CAUTION

It might occur that duplicate documents exist deliberately. Naturally the KM Report does not distinguish between intended and unintended duplicates.

## 2. Background Information

The KM Report identifies duplicate documents on the basis of a distinct value determined for each resource by applying the SHA-1 algorithm.

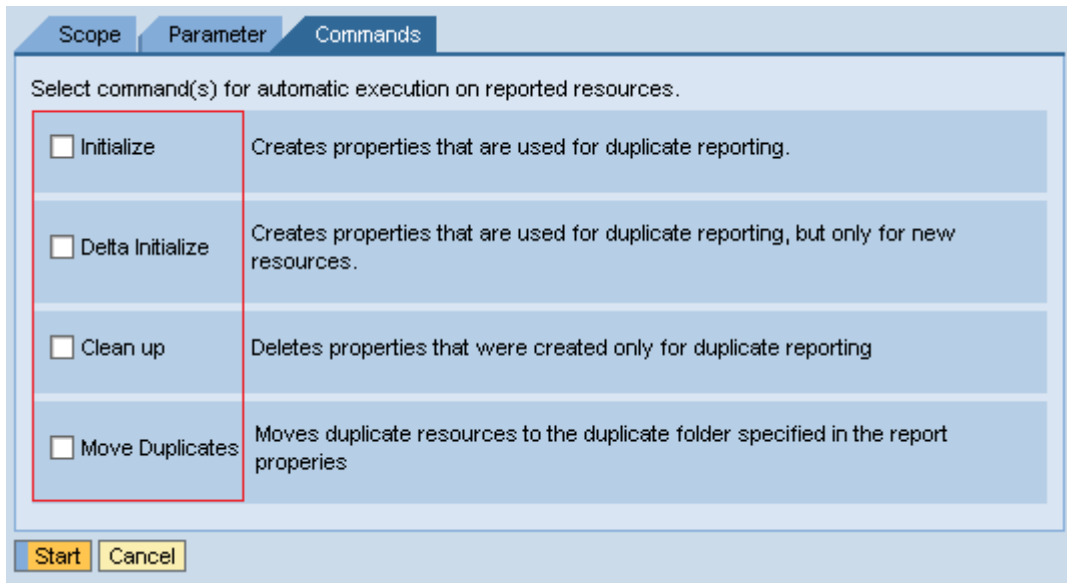
## 3. Prerequisites

- Starting documentation:
  - Get to know KM Reporting: [KM Reports](#).
  - Find out how to deploy Portal Applications: [Deployment of Applications](#).
- Definition: “Duplicate”
  - A KM Resource is considered a duplicate if the **content** of the resource can be considered identical to the content of another resource. Properties or other metadata like permissions are not considered.

### Note

If A duplicates B then B duplicates A and vice versa.

- The KM Report that will be introduced in this How-To-Guide provides different *Commands*:



Scope Parameter **Commands**

Select command(s) for automatic execution on reported resources.

<input type="checkbox"/> Initialize	Creates properties that are used for duplicate reporting.
<input type="checkbox"/> Delta Initialize	Creates properties that are used for duplicate reporting, but only for new resources.
<input type="checkbox"/> Clean up	Deletes properties that were created only for duplicate reporting
<input type="checkbox"/> Move Duplicates	Moves duplicate resources to the duplicate folder specified in the report properties

Start Cancel

**Figure 1: Commands Overview**

- **Initialize** command:  
Run report to calculate a SHA-1 hash **for each** resource and store as application property.
- **Delta Initialize** command:  
Run report to calculate a SHA-1 hash **only for newly added** resources and store as application property.
- **Clean Up** command:  
Run report to delete SHA-1 hash application properties **for each** resource.
- **Move Duplicates** command:  
Run report to **find** documents with identical hash values and **move** them to the prior specified folder.
- If **no command** is selected:  
Run report to **find** documents with identical hash values but do **not move** them.

 **CAUTION**

No more than one of the above-mentioned commands should be selected at a time as they are mutually exclusive. Ignoring this warning might lead in some cases to serious issues.

## 4. Step-by-Step Procedure

This section describes how to obtain, deploy and apply the KM Report called *Duplicate Check Report* in order to **find** duplicate documents and **move** them to a specified folder.

### 4.1 Obtain and Deploy KM Report

#### Note

Please be sure to have the role of a system administrator to be enabled to deploy the PAR file.

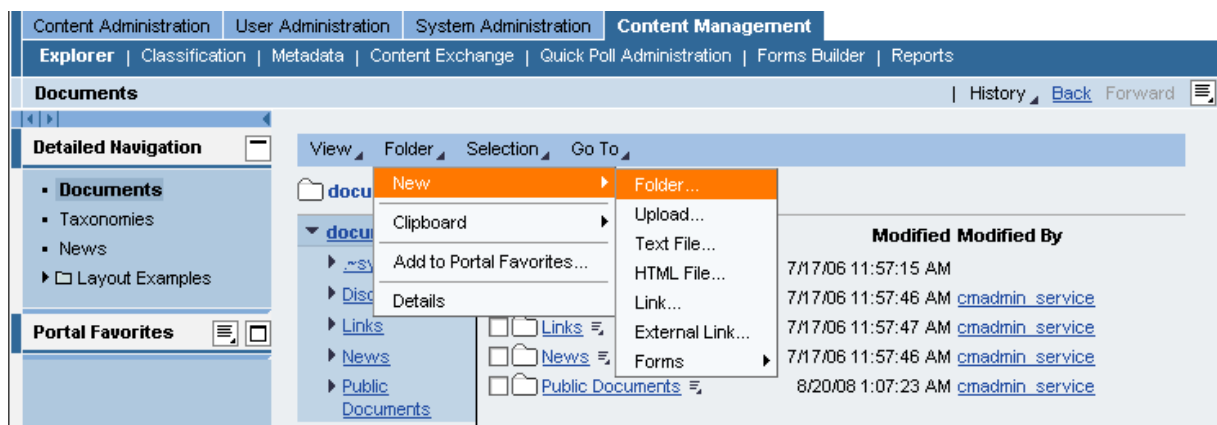
1. Obtain the required KM Report that is available as a ZIP-archive. Please use the “Download” Link provided in the abstract.
2. Deploy the PAR file to the portal.

### 4.2 Prepare Duplicate Check (*Initialize* Command)

#### Note

Please be sure to have the role of a Content Manager to be enabled to follow the following instructions.

1. Navigate to *Content Management* → *Explorer* → *Documents* and create a new folder within this repository.



**Figure 2: Creating a new folder where duplicate documents will be moved to**

2. Name it **Duplicates** and click **Save**.

#### Important

Please notice that it is important to differentiate between the use of capital and small initial letters whenever nominating a folder or entering a folder path.



- Now navigate to *Content Management* → *Reports* → *Tools* where the newly deployed KM Report called *Duplicate Check Report* shows up by now.

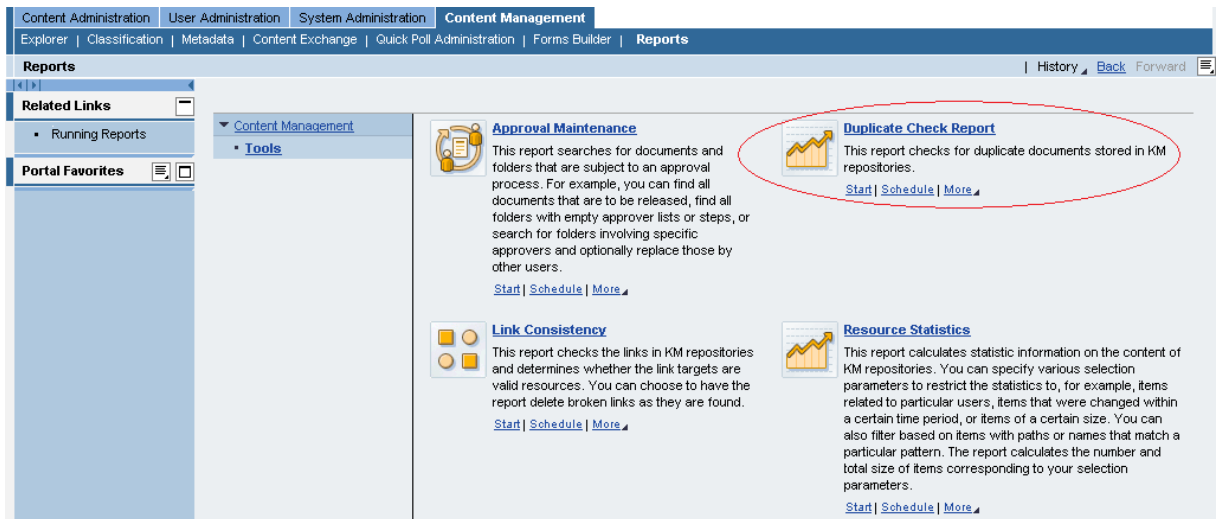


Figure 3: Reports overview

- Click on *Start* and nominate the report.
- Browse* or enter directly the location the report should be employed on, then click *Add*.

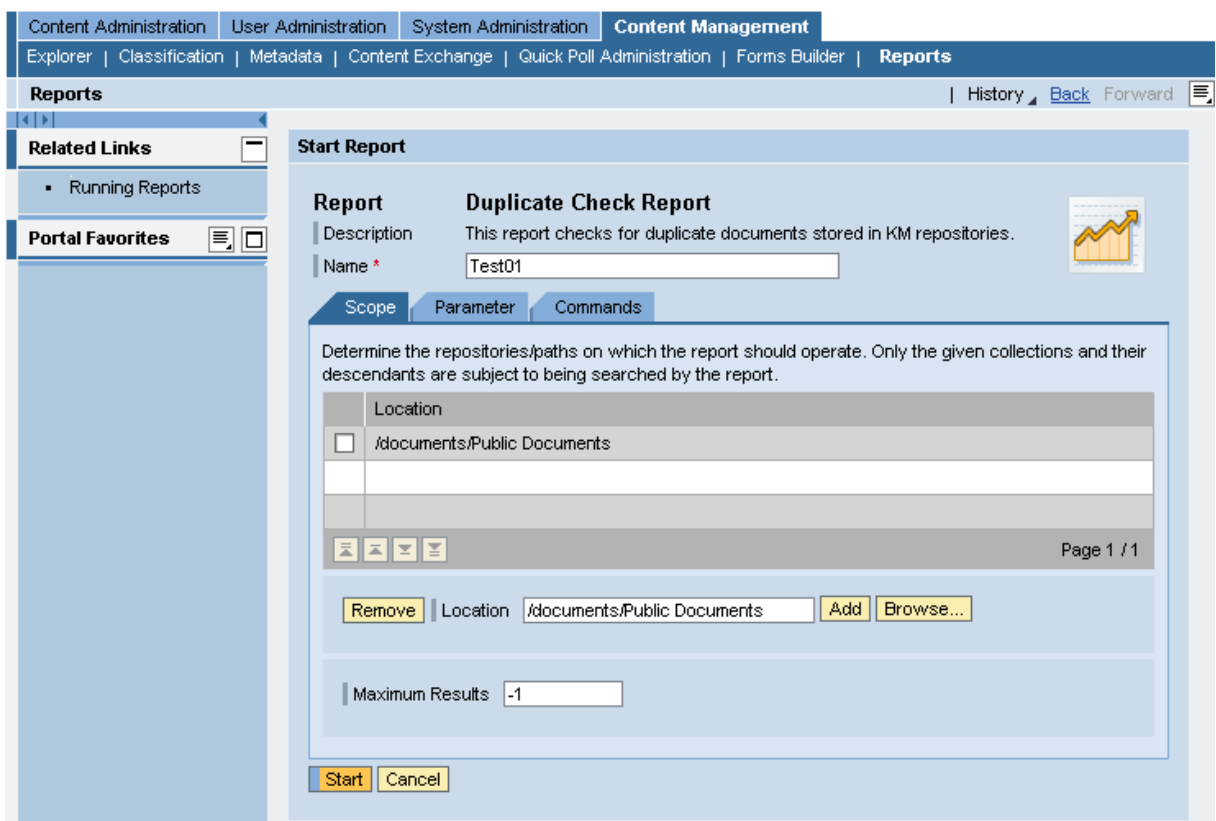


Figure 4: The report's Scope tab

- Switch to *Parameter* tab and specify the folder for duplicates. In case you have followed all previous steps please type in `/documents/Duplicates`.

Important

The specified folder for duplicates needs to be within the same repository than the report's scope selected in the previous step.

Tip

To assure whether the entered path exists, click on *Verify* before going on.

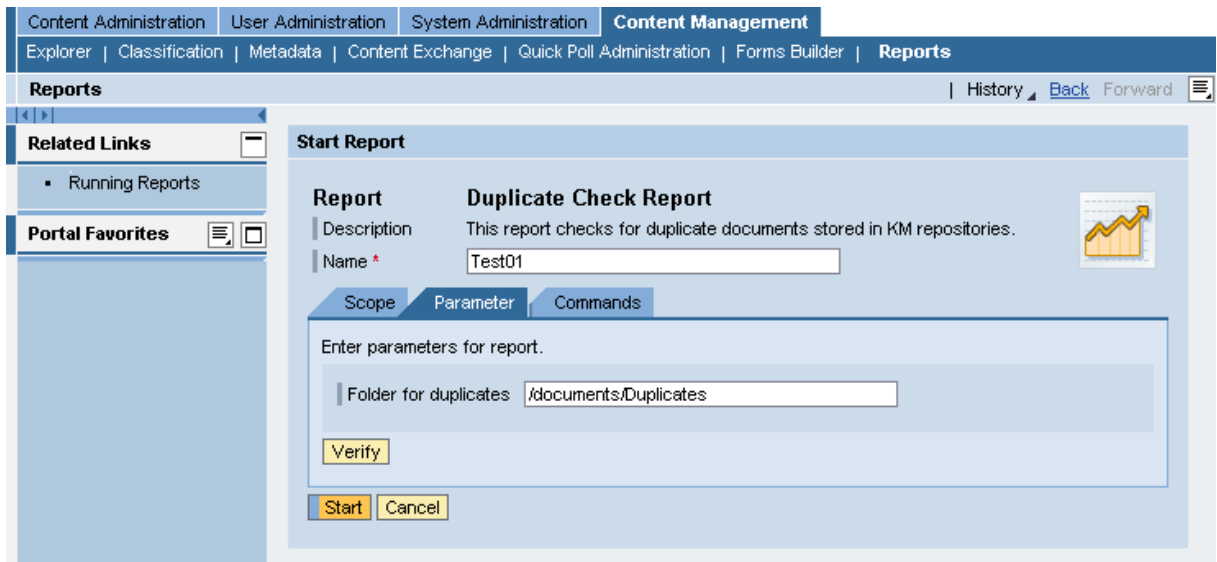


Figure 5: The report's *Parameter* tab

- Now go to *Commands* and select the *Initialize* checkbox.

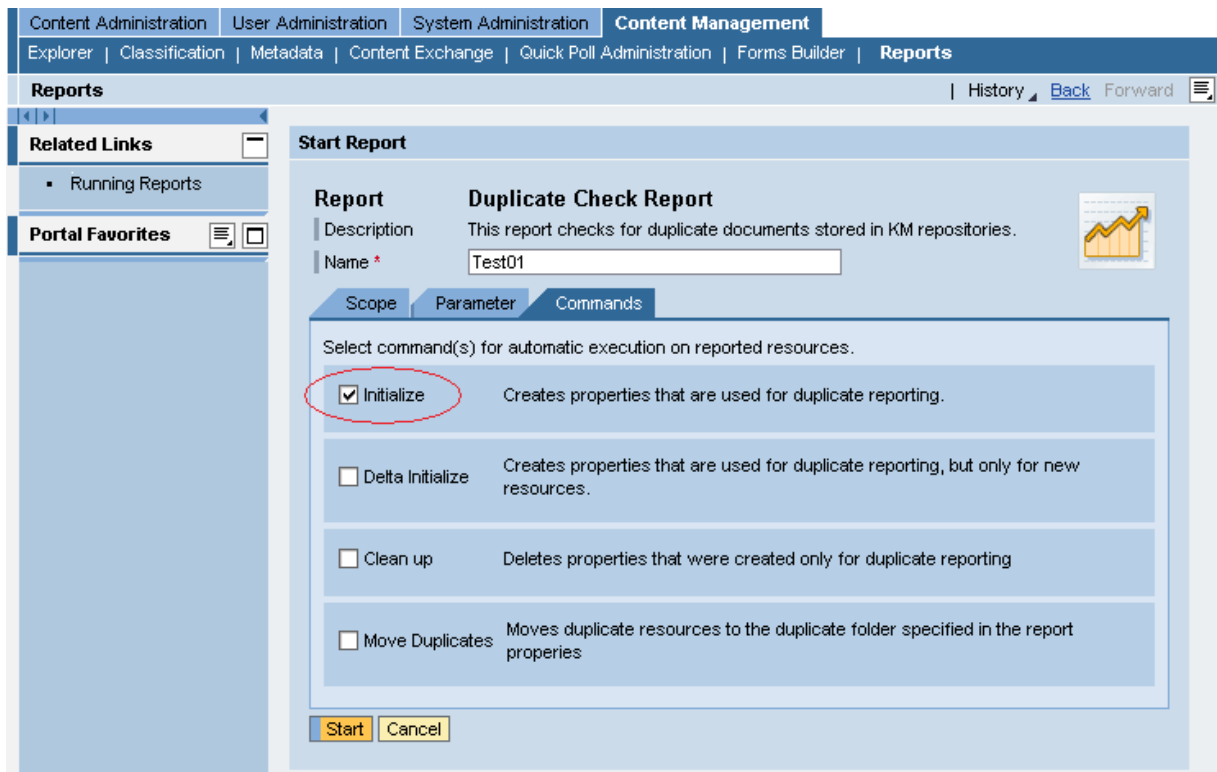


Figure 6: The report's *Commands* tab – Selecting *Initialize* command

8. Click *Start* to execute initialization.

 Note

If the beforehand specified scope of the report already has been initialized once it is recommended to use the *Delta Initialize* Command instead.

### 4.3 Duplicate Check (*Move Duplicates* Command)

 CAUTION

Ensure that the recently started report has completed the initialization before going on.

1. Navigate to *Content Management* → *Reports* → *Tools*.
2. Once again click on *Start* in order to open the *Duplicate Check Report*.
3. Nominate the report.
4. *Browse* or enter directly the location the report should be employed on, then click *Add*.

 CAUTION

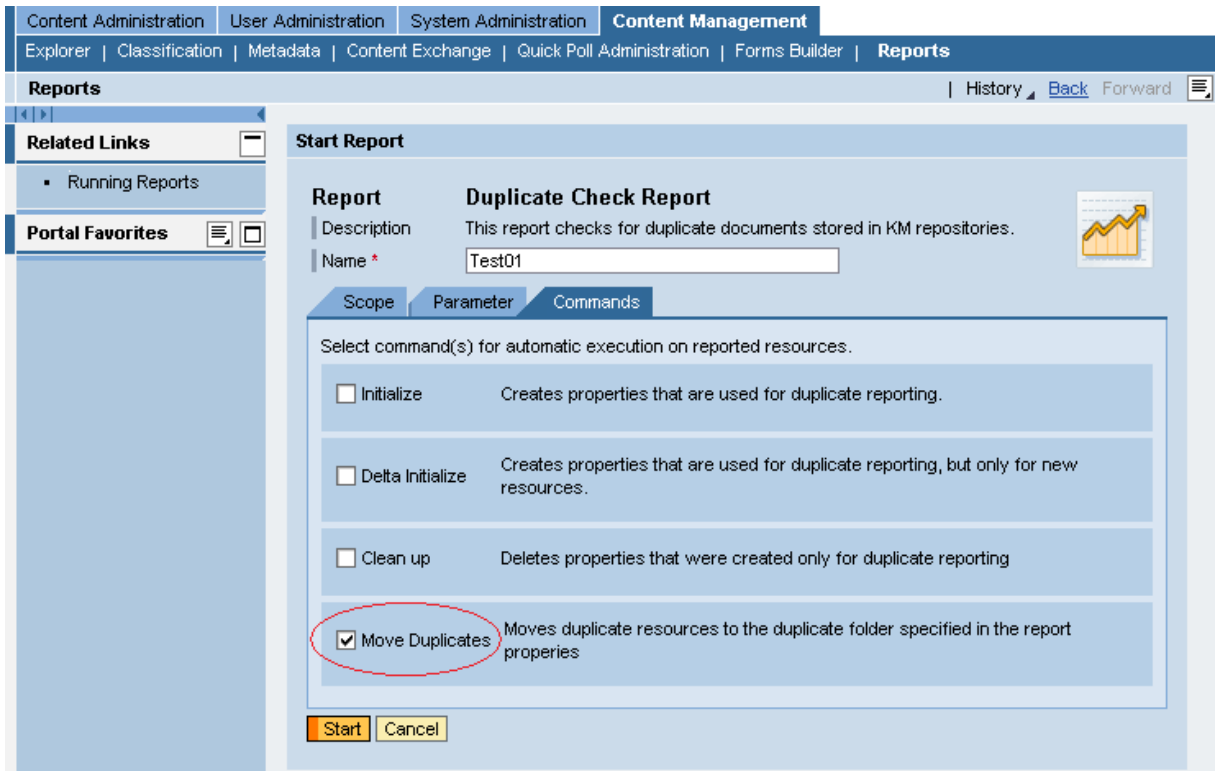
This time do **not** add multiple locations as this might lead to serious issues.

5. Switch to *Parameter* tab and specify the folder where duplicates should be moved to. In case you have followed all previous steps please type in `/documents/Duplicates`.

 Important

The specified folder for duplicates needs to be within the same repository than the report's scope selected in the previous step.

6. Then switch to the *Commands* tab and select the *Move Duplicates* checkbox.



The screenshot shows the SAP Reports interface for the 'Duplicate Check Report'. The 'Commands' tab is active, and the 'Move Duplicates' checkbox is selected and circled in red. The report name is 'Test01'. The description states: 'This report checks for duplicate documents stored in KM repositories.' The 'Commands' section lists four options: 'Initialize', 'Delta Initialize', 'Clean up', and 'Move Duplicates'. The 'Move Duplicates' option is checked and circled in red. The 'Start' button is highlighted in orange.

**Figure 7: The report's *Commands* tab – Selecting *Move Duplicates* command**

7. Click *Start* to move duplicates to the specified folder.
8. Now take a look at the folder where the duplicates – after the report has finished – were moved in order to get to know duplicate documents and decide how to proceed with them.

 **Note**

Instead of moving duplicate documents it is also possible only to detect duplicates by selecting **no command** and clicking *Start*. In order to perceive the list of identified duplicate documents simply take a look at the report itself.

## 5. Appendix

Another type of duplicate check is available [here](#). It enables a check for duplicates for documents at the time when they are uploaded. Its duplicate detection is based on comparing the hash values of documents, too.

[www.sdn.sap.com/irj/sdn/howtoguides](http://www.sdn.sap.com/irj/sdn/howtoguides)