

Sinequa - RegWatch - script to update index

- [Introduction](#)
- [Context](#)
- [Problem with multiple attachments](#)
- [Meta properties changes](#)
- [Main actions of the script](#)
- [Folders](#)
- [Files](#)
- [Configfile.ps1](#)
- [updatesinequa.ps1 \(version 12.04.2024\)](#)

Introduction

This document describes the PowerShell script used to update certain meta-properties of the Regwatch collection.

Context

The Regwatch search interface <https://regwatch-search-eco.solvay.com/> is based on the indexing of the Aodocs RegWatch Database library

The RegWatch Database is an AODocs library for permanent storage of regulatory documents and information collected from regulatory agencies, industry associations, and other sources on chemical, market, and export control regulations.

Problem with multiple attachments

Documents in the Aodocs library can have multiple files attached. When indexing Sinequa creates a record for each file. To have only one item when searching, we create a PowerShell script to modify the data in the index after indexing. The script tags the first record (= first file) of the documents and links the others to the main record.

Meta properties changes

For a document with 3 attached files, In the index we will have after executing the script:

[blocked URL](#)

The sourcestr23 field will contain for the first record the value Main, and the sourcecsv6 field will contain the 2 links (separator= ;) of the 2nd and 3rd files in Aodocs.

With these data change, the search interface will display only documents having the sourcestr23 field equal to Main like this:

[blocked URL](#)

Main actions of the script

```
execute this query: select sourcestr22 from idx_AoDOcs where collection='$Collection' and sourcestr22 <> '' group by sourcestr22 order by sourcestr22 asc;
```

The sourcestr22 field contains doc.url1 (AODocs link to the document)

The result of the query is the list of Aodocs documents.

For each Document, this query is executed: select id,url2 from \$Index where collection='\$Collection' and sourcestr22 = '\$UrlAodocs' order by filename asc;

The first record (<=> first file) is updated:

```
update $Index set sourcestr23 = 'Main',sourcecsv6='$sourcecsv6' where collection = '$Collection' and id='$idmain'
```

sourcecsv6 contains the list of links to the file in Aodocs for the 2nd, 3rd,... records.

Folders

F:\works\regwatch_ECO_V2	Executable, configuration and logs files for Solvay instance
F:\works\regwatch_SCO_V2	Executable, configuration and logs files for Syensqo instance

Files

Configfile.ps1	Configuration file
updatesinequa.ps1	PowerShell executable file
updatesinequa.ps1.log	Log file
updatesinequa.ps1.detailed.Saturday.log	Detailed log file (weekly rotation)

Configfile.ps1

```
$Folder= "F:\works\regwatch_ECO_V2\  
$SinequaAPIUrl = "https://utopia-search.solvay.com/xrest"  
$Collection ="/AoDocs/RegWatch Database/"  
$Index="idx_AoDocs"  
$Credentials="XXXXXXXXXXXX" #wsinequa + pwd generated in postman  
$SinequaUser="powershell"  
$SinequaPassword="XXXXXXXXXX"
```

updatesinequa.ps1 (version 12.04.2024)

Stored in Gitlab: <https://gitlab.solvay.com/cas/sinequa/scripts/regwatch-script-to-update-index>