

Compare and Merge Scripts

Octavian Nadolu, Syncro Soft
octavian.nadolu@oxygenxml.com
@OctavianNadolu

© 2021 Syncro Soft SRL. All rights reserved.

Agenda

- Overview of the **Oxygen Diff Support**
- **File Comparison and Merge** Reports Using Command Line Scripts
- **Directory Comparison and Merge** Reports Using Command Line Scripts
- Generate File and Directory Comparison **Reports in Oxygen**



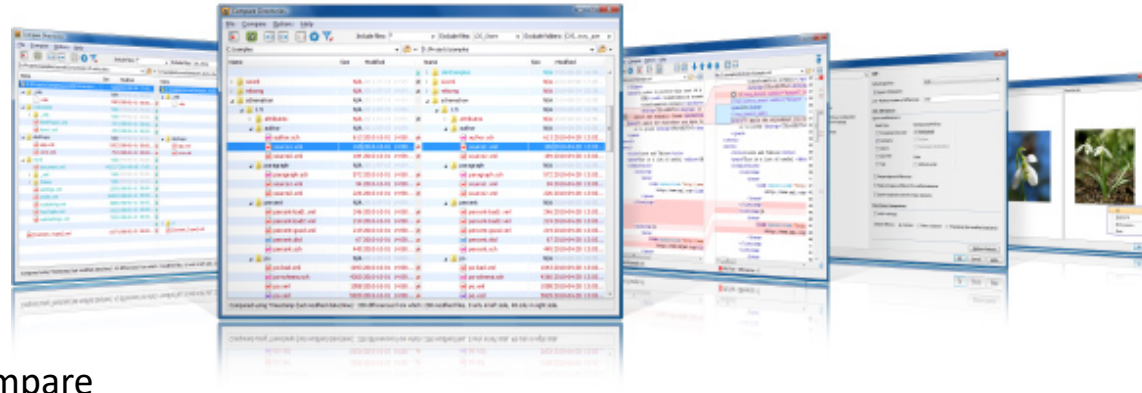
Compare and Merge

- **Compare documents** – identify changes between two versions of the same document
- **Merge documents** - uniting two documents into a single one



Compare and Merge Support in Oxygen

- Compare and Merge Files
 - Text and Visual Author Compare
 - Three-Way Comparisons
 - XML Diff Algorithms
- Compare and Merge Directories
 - Binary, Timestamp, and Content Compare
 - Compare in Archives
 - Streaming Diff Algorithms
- 3-Way Directory Comparison and Merging Changes
- Generating File or Directory Comparison HTML Reports*
- Compare and Merge Files/Directories Command-Line Scripts*

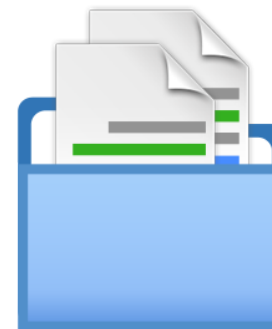


Question: Do you use/intend to use Oxygen Diff, what tools?

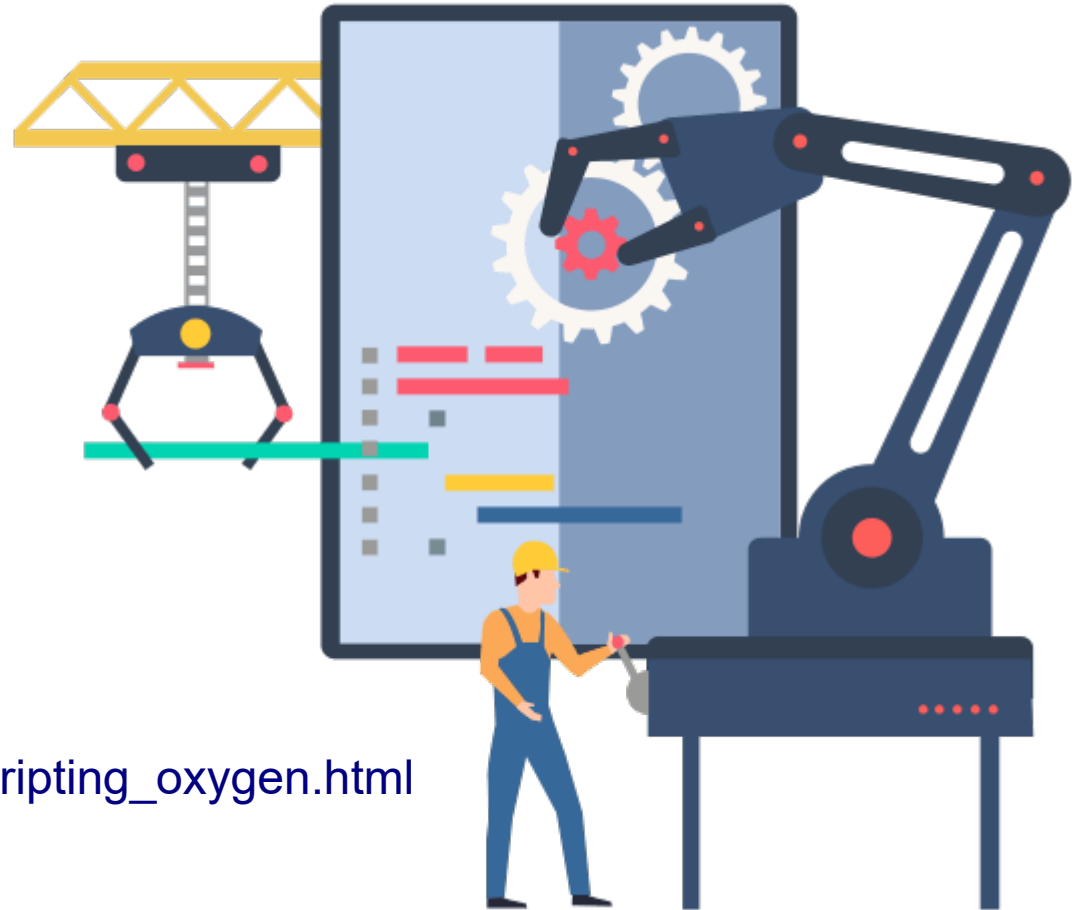
- Compare and Merge Files
- Compare and Merge Directories
- 3-Way Directory Comparison and Merging Changes
- Generate Comparison HTML Reports
- Compare and Merge Scripts



Compare and Merge Scripts



Oxygen XML Scripting



oxygenxml.com/oxygen_scripting.html

oxygenxml.com/doc/ug-editor/topics/scripting_oxygen.html

Oxygen XML Scripting Functions

Automate and run Oxygen utilities from the command-line Interface

- Validation
- Transformation
- Conversion
- Generate Documentation
- Format and Indent
- **Compare and Merge Files or Directories**
- and more

Webinar: Automate XML processing with Oxygen XML Scripting

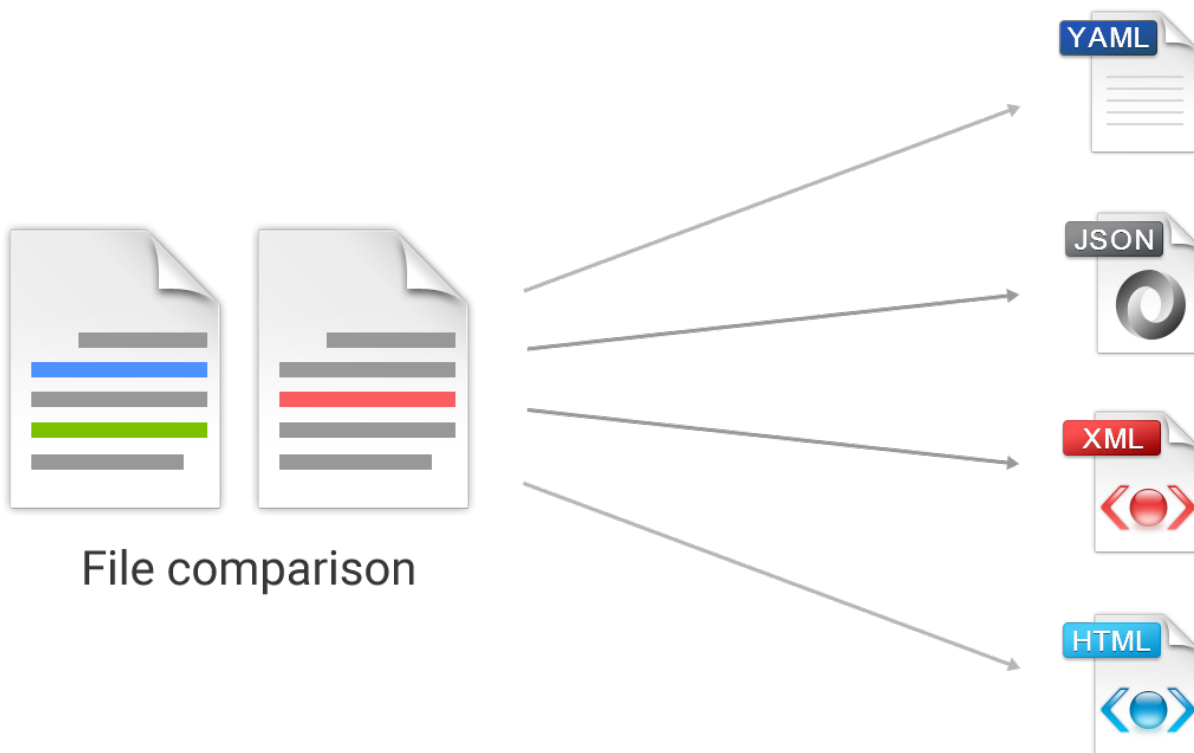
Question: Do you use/intend to use Oxygen Scripting, what tools?

- ❑ Validation
- ❑ Transformation
- ❑ Generate Documentation
- ❑ Compare and Merge
- ❑ Other



Compare and Merge Files Script

Compare and merge files and get the comparison results in various formats



Arguments for the Compare Files Script

firstFilePath - Mandatory argument that specifies the first file path.

secondFilePath - Mandatory argument that specifies the second file path.

baseFilePath - Optional argument that specifies the path of the base file (for 3-way comparisons).

-ct contentType - Specifies the content type of the files to be compared. The option is used to force the file handling to the specific type of file. Otherwise, the file type is auto-detected.

-out outputFormat - Specifies the format of the output. Possible values: *yaml*, *json*, *xml*, *html*, *htm*, *html/inlineCSS*, or *htm/inlineCSS*. Default value = *yaml*.

-outfile outputFile - Specifies the path for an output file to save the comparison results, instead of presenting them in the console.

-help | --help | -h | --h - Displays help text.

Compare Files Script Report

Example of generating a file comparison report in YAML format

```
compareFiles file1 file2
```

Report information:

- Total number of modifications
- Number of modifications by type
- Location of the modifications

```
differences:  
  diffsNo: 7  
  deletions: 4  
  addings: 2  
  modifications: 1  
  diffs:  
  - firstFile:  
    startOffset: 470  
    endOffset: 555  
    secondFile:  
    startOffset: 470  
    endOffset: 470  
    diffType: deleted  
  - firstFile:  
    startOffset: 622  
    endOffset: 625  
    secondFile:  
    startOffset: 537  
    endOffset: 537  
    diffType: deleted
```

Compare Files Script YAML/JSON/XML Report

Can be used to:

- Verify if the documents are modified
- Determine the type of changes between documents
- Process the changes from the documents
- Executed from CI/CD pipelines

Compare Files Script HTML Report

Example of generating a file comparison report in HTML format:

```
compareFiles file1 file2 -out html -outfile outFileName.html
```

Example of generating a file 3-way comparison report in HTML format:

```
compareFiles file1 file2 baseFile -out html -outfile outFileName.html
```

HTML Report for File Comparison

Differences: 5 difference blocks, 8 differences in total

Comparison details by difference blocks: all (5) incoming (3) outgoing (2)

Base file: D:/Sample1/dita-flowers/flowers-base/topics/flowers/gerbera.dita

File 1: D:/Sample1/dita-flowers/flowers-by-John/topics/flowers/gerbera.dita		File 2: D:/Sample1/dita-flowers/flowers-by-Mary/topics/flowers/gerbera.dita
8 is a genus of ornamental plants from the daisy family (Asteraceae). It was named in 9 honor of the German naturalist Traugott Gerber (1710-1743) who travelled extensively in Russia and 10 was a friend of Carl Linnaeus. </p>	+ -	is a genus of ornamental plants from the sunflower family (Asteraceae). It was named in honor of the German naturalist Traugott Gerber. </p>
12 13 <!--Maybe we can add more pictures here.--> 14 15 <p>It has approximately 30 species in the wild, extending to South America, Africa and tropical	+ -	11 <p>It has approximately 30 species in the wild, extending to South America, Africa and tropical
18 also known as Transvaal daisy or Barberton Daisy. </p> 19 <p>Gerbera species bear a large capitulum with striking, two-lipped ray florets in yellow,	- +	14 also known as Transvaal daisy or Barberton Daisy. </p> 15 16 17 18 <p>Gerbera species bear a large capitulum with striking, two-lipped ray florets in yellow,
25 The domesticated <xref keyref="cultivar" format="dita">cultivars</xref> are mostly a result of a	- +	The domesticated <xref format="dita" keyref="cultivar">cultivars</xref> are mostly a result of a

Arguments for the Compare Files Algorithms

-alg comparisonAlg - Specifies the algorithm to be used for the comparison. Possible values: auto, chars, words, lines, syntax_aware, xml_fast, and xml_accurate. Default value = auto.

-als algStrength - Specifies the strength of the algorithm to be used for the comparison. Possible values: low, medium, high, and very_high. Default value = medium.

-iws ignoreWS - If set to true, whitespaces are ignored if differences consist only of whitespaces. Default value = false.

Compare Files Arguments

Example of a comparison that uses the Lines algorithm:

```
compareFiles file1 file2 base -out html -outfile fileName.html -alg lines
```

Arguments for XML-aware Algorithms

- ipi ignorePI** - If set to true, processing instructions are ignored in the comparison. Default value = false.
- icm ignoreComments** - If set to true, comments are ignored in the comparison. Default value = false.
- idt ignoreDocType** - If set to true, DOCTYPE sections are ignored in the comparison. Default value = false.
- itn ignoreText** - If set to true, text content is ignored in the comparison. Default value = false.
- ins ignoreNS** - If set to true, namespaces are ignored in the comparison. Default value = false.
- ind ignoreNSDecl** - If set to true, namespace declarations are ignored in the comparison. Default value = false.
- inp ignorePrefixes** - If set to true, prefixes are ignored in the comparison. Default value = false.
- iao ignoreAttrOrder** - If set to true, the order of attributes is ignored in the comparison. Default value = false.
- iee ignoreExpStateForEmptyElems** - If set to true, the expansion state for empty elements is ignored in the comparison. Default value = false.
- enx XPathExprToExcludeNodes** - Specifies an XPath expression to exclude certain nodes from the comparison.

Compare Files Ignore Options

Example of a comparison that ignores whitespaces, comments, and attributes order:

```
sh compareFiles.sh file1 file2 baseFile -iws -icm -iao -out html -outfile  
outFileName.html
```

Merge Files Arguments

-merge mergeOperation - If set to true, a merge operation is invoked after the comparison. Default value = true.

-mergeout outputDirPathForMerge -Specifies the output directory path for the merged file (instead of directly affecting the first compared file).

- The merge operation is similar to the same process in any versioning system
- If conflicting changes are detected, the merge operation is aborted and the first file remains unchanged
- After the comparison and merge, a report is created that provides some details about the changes that were made

Merge Files Script

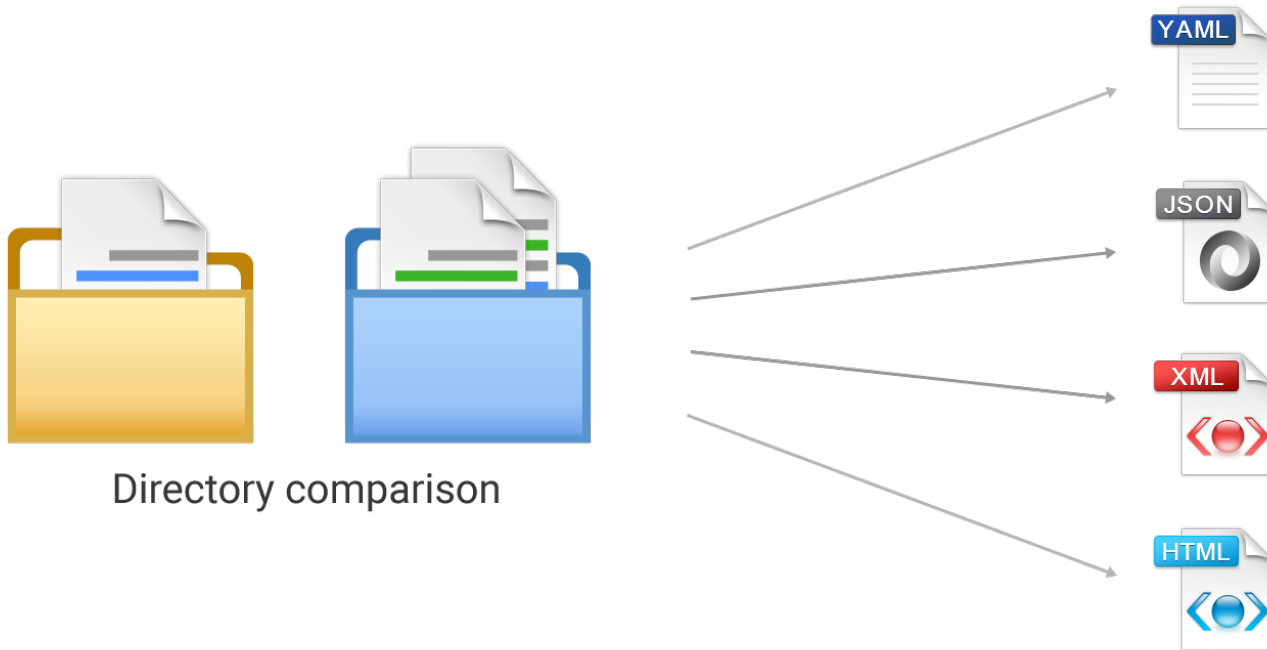
Example of merging the changes between two files:

```
compareFiles file1 file2 baseFile -mergeout destFolder
```

```
mergeReport:  
  operationStatus: "OPERATION COMPLETED"  
  changesFound: 8  
  modificationsPerformed: 8
```

Compare and Merge Directories Script

Compare and merge directories and get the comparison results in various formats



Arguments for the Compare Directories Script

firstDirPath - Mandatory argument that specifies the first directory path.

secondDirPath - Mandatory argument that specifies the second directory path.

baseDirPath - Optional argument that specifies the path of the base directory (for 3-way comparisons).

-out outputFormat - Specifies the format of the output. Possible values: yaml/grouped, yaml/raw, json/grouped, json/raw, xml/grouped, xml/raw, html, html/ifcr, htm, or htm/ifcr. Default value = yaml/grouped.

-outfile outputFile - Specifies the path for an output file to save the comparison results, instead of presenting them in the console. The content of the output file is formatted according to the -out argument.

-help | --help | -h | --h - Displays help text.

Compare Directories Script Report

Example of generating a directories comparison report in YAML format

```
compareDirs dir1 dir2
```

Report information:

- Total number of modifications
- Number of modifications by type
- Modified files

```
differences:
  diffsNo: 14
  onlyFoundInFolder1:
    filesNo: 2
    files:
      - publishing\flowers\README.txt
      - topics\test
  onlyFoundInFolder2:
    filesNo: 2
    files:
      - publishing\flowers\resources\image
      - publishing\flowers\READ-ME.txt
  modified:
    filesNo: 10
    files:
      - concepts\autumnFlowers.dita
      - concepts\glossaryGenus.dita
      - concepts\glossaryPanicle.dita
      - images\Gerbera.jpg
      - tasks\gardenPreparation.dita
      - topics\flowers\chrysanthemum.dita
      - topics\flowers\gerbera.dita
      - topics\flowers\gerbera.json
      - topics\flowers\snowdrop.dita
      - topics\introduction.dita
```


Compare Directories Script Output Format

-out outputFormat - Specifies the format of the output. Possible values: yaml/grouped, yaml/raw, json/grouped, json/raw, xml/grouped, xml/raw, html, html/ifcr, htm, or htm/ifcr. Default value = yaml/grouped.

- If you choose to save/redirect the console output to a file, this argument establishes the type of the output file and its content is formatted accordingly. If you skip specifying the "/grouped" or the "/raw" qualifiers, "/grouped" takes precedence.
- If you choose the html or htm output format, it is recommended that you also choose to save/redirect the console to the specified HTML file to view the comparison result in your preferred browser.
- The "/ifcr" qualifier for the html or htm values is considered only if the -outfile argument is also present. IFCR is an acronym for Include File Comparison Reports and it means that, along with generating the directory comparison report, separate file comparison reports will be generated for all modified file pairs. These reports are available through links from the main report and are saved to a specific directory based on the value provided by the outfile argument. It will have the same parent directory and the same name as the outputFile plus -OXY-FC-REPORTS added to the end of its name.
- The html value, as well as the grouped, raw, or ifcr qualifiers, are not considered if the -merge argument is present.

Compare Directories Script HTML Report

Example of generating a directory comparison report in HTML format:

```
compareDirs dir1 dir2 -out html -outfile results.html
```

Example of generating a 3-way directory comparison report in HTML format:

```
compareDirs dir1 dir2 baseDir -out html -outfile results.html
```

Example of generating a 3-way directory comparison including files reports:

```
compareDirs dir1 dir2 baseDir -out html/ifcr -outfile results.html
```

HTML Report for Directory Comparison

Differences: 13

Comparison details: all differences (13) outgoing (5) incoming (5) conflicts (3)

Base folder: D:/Sample1/dita-flowers/flowers-base/

Folder 1: D:/Sample1/dita-flowers/flowers-by-John/ Folder 2: D:/Sample1/dita-flowers/flowers-by-Mary/

File name	Size	Modified		File name	Size	Modified
concepts/autumnFlowers.dita	1151	2021-07-06 01:49:12	* ×	concepts/autumnFlowers.dita	1143	2021-07-16 06:52:21
concepts/glossaryGenus.dita	571	2021-07-16 06:54:17	*	concepts/glossaryGenus.dita	577	2021-07-06 01:49:12
concepts/glossaryPanicle.dita	483	2021-07-15 06:53:34	*	concepts/glossaryPanicle.dita	495	2021-07-06 01:49:12
images/Gerbera.jpg	10134	2021-07-06 01:49:12	*	images/Gerbera.jpg	22776	2021-07-16 05:55:25
				+ publishing/flowers/resources/images/flower_logo.png	6178	2021-07-06 01:49:12
				+ publishing/flowers/READ-ME.txt	127	2021-07-06 01:49:12
publishing/flowers/README.txt	127	2021-07-06 01:49:12	-			
tasks/gardenPreparation.dita	2275	2021-07-06 01:49:12	*	tasks/gardenPreparation.dita	2291	2021-07-15 12:21:02
topics/flowers/chrysanthemum.dita	2949	2021-07-15 07:48:25	*	topics/flowers/chrysanthemum.dita	2932	2021-07-06 01:49:12
topics/flowers/gerbera.dita	2432	2021-07-16 06:18:28	* ×	topics/flowers/gerbera.dita	2456	2021-07-16 06:25:13
topics/flowers/snowdrop.dita	2883	2021-07-15 13:57:59	*	topics/flowers/snowdrop.dita	2806	2021-07-06 01:49:12
topics/test/		2021-07-15 12:36:56	+			
topics/introduction.dita	770	2021-07-16 06:58:21	* ×	topics/introduction.dita	758	2021-07-15 12:12:46

Arguments for Compare Directories Filter

- ia includeArchives** - If set to true, files from archives are included in the comparison. Default value = false.
- if includeFilesFilter** - Use this argument to only include files that match the specified pattern in the comparison (e.g. .xml, .json). Default value = *.
- ef excludeFilesFilter** - Use this argument to exclude files that match the specified pattern from the comparison (e.g. *.jpg).
- ed excludeSubdirsFilter** - Use this argument to exclude sub-directories that match the specified pattern from the comparison (e.g. .svn, _svn, .git).

Arguments for the Compare Directories Script

Example of generating a directory comparison excluding JPEG files

```
compareDirs dir1 dir2 -out html -outfile results.html -ef *.jpg
```

Example of generating a directory comparison including just Dita files

```
compareDirs dir1 dir2 -out html -outfile results.html -if *.dita
```

Arguments for Compare Directories Algorithms

-cm comparisonMode - Specifies the comparison mode. There are three modes available: content, binary, and timestamp. Default value = content.

-alg comparisonAlg - Specifies the algorithm to be used for the comparison. Possible values: auto, chars, words, lines, syntax_aware, xml_fast, and xml_accurate. Default value = auto.

-als algStrength - Specifies the strength of the algorithm to be used for the comparison. Possible values: low, medium, high, and very_high. Default value = medium.

-iws ignoreWS - If set to true, whitespaces are ignored if differences consist only of whitespaces. Default value = false.

Arguments XML-aware Algorithms

- ipi ignorePI** - If set to true, processing instructions are ignored in the comparison. Default value = false.
- icm ignoreComments** - If set to true, comments are ignored in the comparison. Default value = false.
- idt ignoreDocType** - If set to true, DOCTYPE sections are ignored in the comparison. Default value = false.
- itn ignoreText** - If set to true, text content is ignored in the comparison. Default value = false.
- ins ignoreNS** - If set to true, namespaces are ignored in the comparison. Default value = false.
- ind ignoreNSDecl** - If set to true, namespace declarations are ignored in the comparison. Default value = false.
- inp ignorePrefixes** - If set to true, prefixes are ignored in the comparison. Default value = false.
- iao ignoreAttrOrder** - If set to true, the order of attributes is ignored in the comparison. Default value = false.
- iee ignoreExpStateForEmptyElems** - If set to true, the expansion state for empty elements is ignored in the comparison. Default value = false.

Merge Directories Arguments

- merge mergeOperation** - If set to true, a merge operation is invoked after the comparison. Default value = true.
- mergeout outputDirPathForMerge** - Specifies the output directory path for the merge operation.
- The merge operation is similar to the same process in any versioning system. Following the comparison between the first and second directories (relative to the base folder), all the differences of the type incoming are considered and the content of the first directory is updated accordingly.
- For conflicting files that could not be merged, the left file content is kept.
- After the comparison, a report is created that provides details about the changes that were made.

Merge Directories Script

Example of merging the changes between two directories:

```
compareDirs dir1 dir2 baseDir -mergeout destFolder
```

```
mergeReport:
operationStatus: "OPERATION COMPLETED"
added:
  filesNo: 2
  files:
    - publishing\flowers\resources\images
    - publishing\flowers\READ-ME.txt
deleted:
  filesNo: 1
  files:
    - publishing\flowers\README.txt
updated:
  filesNo: 2
  files:
    - images\Gerbera.jpg
    - tasks\gardenPreparation.dita
merged:
  filesNo: 1
  files:
    - topics\flowers\gerbera.dita
conflicts:
  filesNo: 3
  files:
    - concepts\autumnFlowers.dita
    - topics\flowers\gerbera.json
    - topics\introduction.dita
```

Save HTML Comparison Report in Oxygen

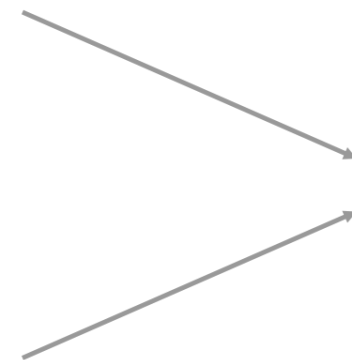
- **Compare Files** – save result in HTML format
- **Compare Directories** – save result in HTML format
- **Generate HTML Report** for Directory Comparison



Diff Files

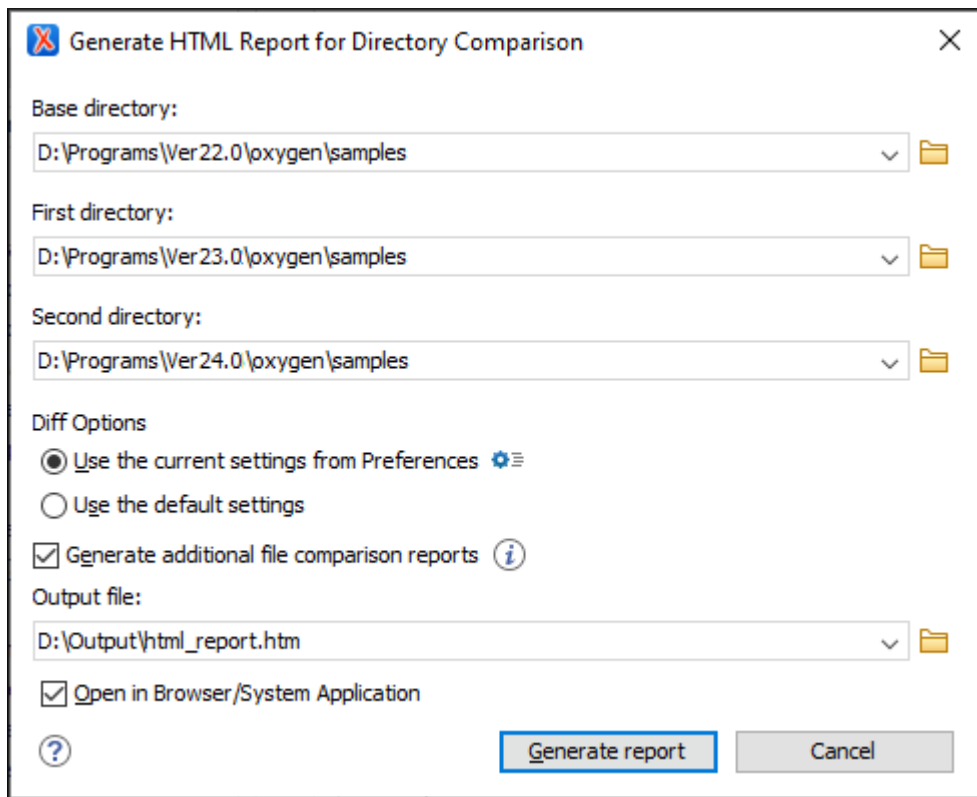


Diff Directories





HTML reports

Generate HTML Report for Directory Comparison



The screenshot shows a dialog box titled "Generate HTML Report for Directory Comparison" with a close button (X) in the top right corner. The dialog contains the following fields and options:

- Base directory:** A text box containing "D:\Programs\Ver22.0\oxygen\samples" with a folder icon on the right.
- First directory:** A text box containing "D:\Programs\Ver23.0\oxygen\samples" with a folder icon on the right.
- Second directory:** A text box containing "D:\Programs\Ver24.0\oxygen\samples" with a folder icon on the right.
- Diff Options:**
 - Use the current settings from Preferences 
 - Use the default settings
- Generate additional file comparison reports 
- Output file:** A text box containing "D:\Output\html_report.htm" with a folder icon on the right.
- Open in Browser/System Application

At the bottom of the dialog, there is a question mark icon on the left, a "Generate report" button in the center, and a "Cancel" button on the right.

Conclusion

- **Complete support** for File and Directory Comparison
- **Scripts** to generate File and Directory **comparison reports**
- **Scripts to merge** Files and Directories
- **Options to filter** the differences
- **Multiple algorithms** to compare different type of files

Future Plans

- Improve Compare Script Algorithms
- Provide Other Compare Reports
- JSON Schema Documentation Script
- OpenAPI Documentation Script
- Other (feedback is welcome)



Question: What features are the most important for you?

- ❑ Improve Compare Script Algorithms
- ❑ Provide Other Compare Reports
- ❑ JSON/OpenAPI Schema Documentation Script
- ❑ Apply Validation Scenario Script
- ❑ Other (feedback is welcome)



Resources

- oxygenxml.com/demo/generating_file_or_directory_comparison_reports_in_oxygen.html
- oxygenxml.com/demo/generating_file_comparison_reports_using_command_line.html
- oxygenxml.com/demo/generating_directory_comparison_reports_using_command_line.html
- oxygenxml.com/demo/Merge_Directories.html
- oxygenxml.com/demo/diff_eclipse.html
- oxygenxml.com/demo/three_way_diff.html
- oxygenxml.com/demo/XML_Diff_Merge.html



Questions?

Octavian Nadolu
Product Manager at Syncro Soft

octavian.nadolu@oxygenxml.com

Twitter: [@OctavianNadolu](https://twitter.com/OctavianNadolu)

LinkedIn: [octaviannadolu](https://www.linkedin.com/in/octaviannadolu)