Storing document images of classification and extraction training sets externally from the TotalAgility database

307765

Question / Problem:
Is it possible to store the document images of classification and extraction training sets externally from the TotalAgility database?

Answer / Solution:
Yes, it is possible to keep training sets out of the database.

Potential Benefits
In some instances, it may be desirable to keep training documents on disk and out of the KTA database. For example, in situations where there is extra sensitivity toward the documents, the documents might be stored on a file system only accessible to the specific developers working on the Transformation project. With this data not in the database, it would also not be included in an exported package of the project and thus neither would this it be included when deploying a project from one environment to another.

When the training sets are on disk, this also eliminates the need for them to be downloaded from the database when the project is opened in Transformation Designer, as well as eliminating the need to save them back to the database when the project is saved. For particularly large training sets, this may be a noticeable reduction in project load and save time.

Consideration: Classification Online Learning
Generally training set data is not needed at runtime. However, the exception is when the Classification project has Online Learning enabled. For these projects, these steps should not be used, and the training sets should remain in the database. Attempting to use an external training set with Classification Online Learning enabled may lead to an error such as this in Transformation Server:

*Failed to perform KTT Online Learning on folder f5b24ae1-766f-47ce-9374-a97600ff40af: Kofax.CEBPM.CPUServer.Common.CPUServerException: Error from ExtractionProcess.exe: Failed to analyze the
Consideration: Access to External Training Sets

To be able to train the project or use any functionality related to training, the person opening the project must have access to the path that contains the training sets. It is important to understand that the path will be used as provided, so if a mapped drive were used, and then a new user opens the project without that drive mapped, it will fail. A persistent UNC path available to all project developers would be recommended.

Moving Training Set From Database to Disk

1. Open a project in Transformation Designer
2. Open the temp location where the project is on disk (for convenience, right click on a training document and click “Open in Windows Explorer”)
3. Copy the entire desired folder (ClassificationTraining and/or ExtractionTraining) to a safe location where they will persist
   a. Understand that this will be connected by the path as provided, so if a mapped drive were used, then a new user opens the project without that drive mapped, it will fail. A persistent UNC path available to all project developers would be recommended.
4. Open the copied folders as new test sets from their new locations.
5. On the newly created test sets, right click and choose “Use as Classification Training Set” or “Use as Extraction Training Set” as appropriate. This will swap the original training sets to be normal test sets, and the new test sets become the training sets.
6. The original training sets (which are now test sets) still live in the project folder and will still be synchronized to the database, so for each, click on the <All Documents> subset, delete all documents. This will remove them from disk. Once the project is saved this will sync a blank training set into the database.
7. Now the project will have the training sets outside of the database. To be able to train the project or use any functionality related to training, the person opening the project must have access to the path that contains the training sets.

Moving Training Set From Disk to Database

1. Open a project in Transformation Designer
2. Open the current user’s temp folder by navigating to “%temp%” in Windows Explorer (which will expand to a path such as “C:\Users\UserName\AppData\Local\Temp”)
3. The currently open project will be in a subfolder with a temporary name such as “rkvc0pxu.w21”. It will most likely be the most recently modified folder, and should contain ClassificationTraining and ExtractionTraining folders.
4. Copy the contents of the external training set into the ClassificationTraining and/or ExtractionTraining folders as appropriate.
5. After the copy completes, open the ClassificationTraining and/or ExtractionTraining folders as a tests set in Transformation Designer.
6. Right click on the newly opened test set and click “Use as Classification Training Set” or “Use as Extraction Training Set” as appropriate, and click OK if prompted to confirm.
7. Train the project (Process > Training > Classification or Extraction)
8. Release the project, which will now include the training set when saving back to the database.

---

**Applies to:**

<table>
<thead>
<tr>
<th>Product</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>KTA</td>
<td>7.0+</td>
</tr>
</tbody>
</table>

---

Author: Stephen Klancher