

How t5memory handles segment saving and matching

This article explains, what segments overwrite others under what conditions.

How segments are saved

Each source segment, that differs from all other source segments, a new source entry is created in the btree of t5memory.

If source is identical (100% match), but target differs regarding content or attributes (except attribute creation date), a new target variant for the source is created.

If a target variant exists and is identical with what is being saved regarding target content and all attributes except creation date, the newer creation date is used.

Attributes are always considered as target attributes, not as source attributes.

This means for example, if 2 segments are identical in source and target and only the creator differs, 2 variants are saved.

If in an imported TMX for a segment creationId and changeld exist, the changeld is saved as creator of the variant.

If in an imported TMX for a segment creationDate and changeDate exist, the changeDate is saved as creationDate of the variant.

The reason for a non existence of a changeld or changeDate in t5memory is, that without having the actual status of the segment at time of creation, distinguishing between creationDate and creationId on the one hand and changeld and changeDate on the other hand is meaningless. Therefore t5memory saves all variants - no information is lost in difference to other TM tools - but there only exists a creator and a creationDate of the variant.

What attributes exist

- author (creator)
- creationDate/timestamp
- documentName: Name of the file from what the segment stems from
- context: Currently the internal DB-segmentId of the segment in translate5 (not the segment number in the task).
Will be changed within 2024 to the segment number in the task, where the segment stems from. Or (if exists) the value of the rename attribute of the translated xiff file (what is drawn from the string ID in JSON or resource files, if filter in translate5 is configured correctly).
- Additional info: Currently unused. In future development steps can be used for any kind of custom attributes.

How 100%-matches and higher are calculated

- If the source is identical, you get a 100% match
- If you use content protection and the protected pattern in the source is the same as in the source of the TM match and the rest of the segment is also the same, you also get a 100% match
- If you have a 100% match and the documentName is also identical, you have a 101% match
- If you have a repetition inside the same task (regardless of how many documents it contains) with all tags of the same type in the same place (type means "opening" tag, "closing" tag or "placeholder" tag) you have a 102% match
- If you have a 100% match with documentName identical and context identical, you have a 103% match.

How 100%-matches are used

- If the config `runtimeOptions.LanguageResources.opentm2.showMultiple100PercentMatches` is activated, all 100%-matches within a TM are shown. Else only the latest within a TM is shown.
- If there are multiple matches of the same match rate (be it from the same TM or different TMs) always the latest with the highest match rate equal or above 100% is used.