Metadata
What your translation systems and files know about you

tekom/tcwold 2019

Angelika Zerfaß

zerfass@zaac.de

Your opinion is important to us! Please tell us what you thought of the lecture. We look forward to your feedback under

http://lt02.honestly.de

or scan the QR code

The feedback tool will be available even after the conference!

Metadata

○ Metadata types and where metadata is saved

○ Uses of metadata

○ What else your tool knows about you

METADATA TYPES

Metadata Types

○ System data
  ○ Automatically saved by the system you work in
○ Tool-specific data
  ○ Metadata fields provided by the translation system
○ Customized data
  ○ Custom-created data fields
○ Process data
  ○ What has happened to the data
  ○ Where does the data come from
  ○ What is the context

System Data

○ User name

- ○ For the user who created/saved a translation unit to a TM or an entry to a term base.
- ○ For the user who used/changed the entry.
- ○ In association with the history (changes) of a project.
- ○ In association with the history (changes) of a segment or term.
- ○ In association with the role (translator, reviewer…) a user had when working on the segments.

○ Date and time when an entry was created, used or changed.
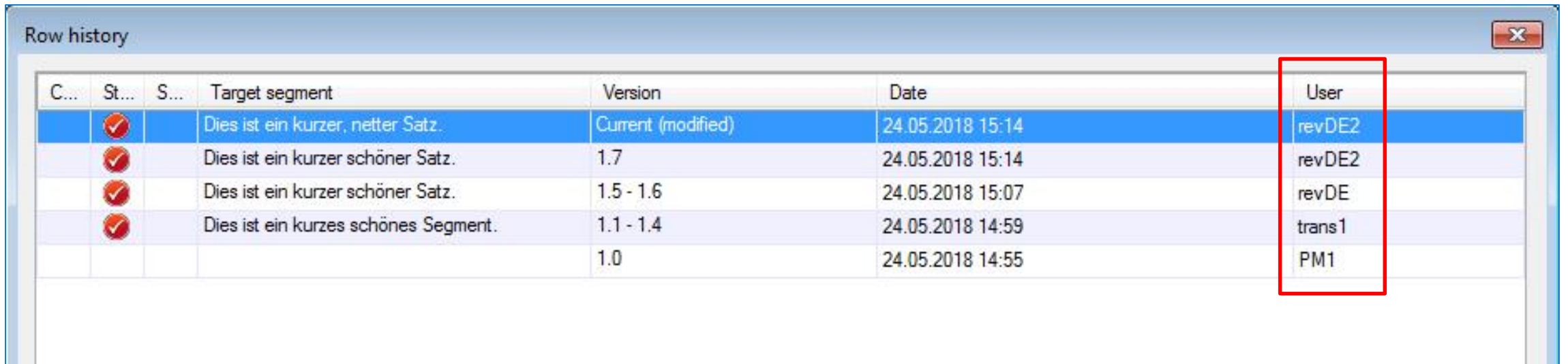
○ Counters on how often a segment was used.

Example: project history

| Project history | Notification history |

| Who | | Document | | Language | All languages ▼ | Refresh history | Reset |

Who	When	Document	Action	Details
PM1	24.05.2018 15:08:53	Demo 1.docx (ger-DE)	snapshot created	minor version: **v6**
revDE	24.05.2018 15:07:36	Demo 1.docx (ger-DE)	delivered	new workflow status: **Review 2 (not started)**
revDE	24.05.2018 15:05:36	Demo 1.docx (ger-DE)	started working on project	
PM1	24.05.2018 15:03:07	Demo 1.docx (ger-DE)	snapshot created	minor version: **v4**
trans1	24.05.2018 15:01:43	Demo 1.docx (ger-DE)	delivered	new workflow status: **Review 1 (not started)**
trans1	24.05.2018 14:57:48	Demo 1.docx (ger-DE)	started working on project	
PM1	24.05.2018 14:56:37	Demo 2.docx (ger-DE)	assignment changed	role: **Translator**, new user: **trans1**, deadline: **24.05.2018 14:56:31**
PM1	24.05.2018 14:56:37	Demo 2.docx (ger-DE)	assignment changed	role: **Reviewer1**, new user: **revDE**, deadline: **24.05.2018 14:56:35**
PM1	24.05.2018 14:56:37	Demo 2.docx (ger-DE)	assignment changed	role: **Reviewer2**, new user: **revDE2**, deadline: **24.05.2018 14:56:38**
PM1	24.05.2018 14:56:37	Demo 1.docx (ger-DE)	assignment changed	role: **Translator**, new user: **trans1**, deadline: **24.05.2018 14:56:29**
PM1	24.05.2018 14:56:37	Demo 1.docx (ger-DE)	assignment changed	role: **Reviewer1**, new user: **revDE**, deadline: **24.05.2018 14:56:33**
PM1	24.05.2018 14:56:37	Demo 1.docx (ger-DE)	assignment changed	role: **Reviewer2**, new user: **revDE2**, deadline: **24.05.2018 14:56:36**
PM1	24.05.2018 14:55:31	Demo 2.docx (ger-DE)	original import	
PM1	24.05.2018 14:55:31	Demo 1.docx (ger-DE)	original import	

Segment History

User name, role and time stamp in the bilingual file

```
translatorcommitusername="trans1" translatorcommittimestamp="2018-05-24T12:59:43Z"
reviewer1commitusername="revDE" reviewer1committimestamp="2018-05-24T13:35:16Z"
reviewer2commitusername="revDE2" reviewer2committimestamp="2018-05-24T13:36:46Z"
```

Example: term entry history

Entry history			
Modified at	**User name**	**Change**	**Value**
5/25/2018 11:43:53 AM	pmaz	+field (eng-US / screen / Abteilung)	Vertrieb
5/25/2018 11:43:53 AM	pmaz	@field (eng-US / screen / Abteilung)	Einkauf => Vertrieb
5/25/2018 11:43:53 AM	pmaz	@term (ger-DE)	Blds. => Bs.
5/25/2018 11:43:53 AM	pmaz	@field (ger-DE / Blds. / Case sensitivity)	Near => Strict
5/25/2018 11:43:53 AM	pmaz	@field (ger-DE / Bildschirm / Abteilung)	Einkauf => Vertrieb
7/1/2016 10:33:44 AM	admin	+field (ger-DE / Blds. / Abteilung)	Einkauf
7/1/2016 10:33:44 AM	admin	+field (ger-DE / Bildschirm / Abteilung)	Einkauf

Example: usage counter

Usage counter: when a suggestion from the TM was used as it is (not changed). Can be used to delete segments from a TM that were not (often) used.

Using Metadata (names)

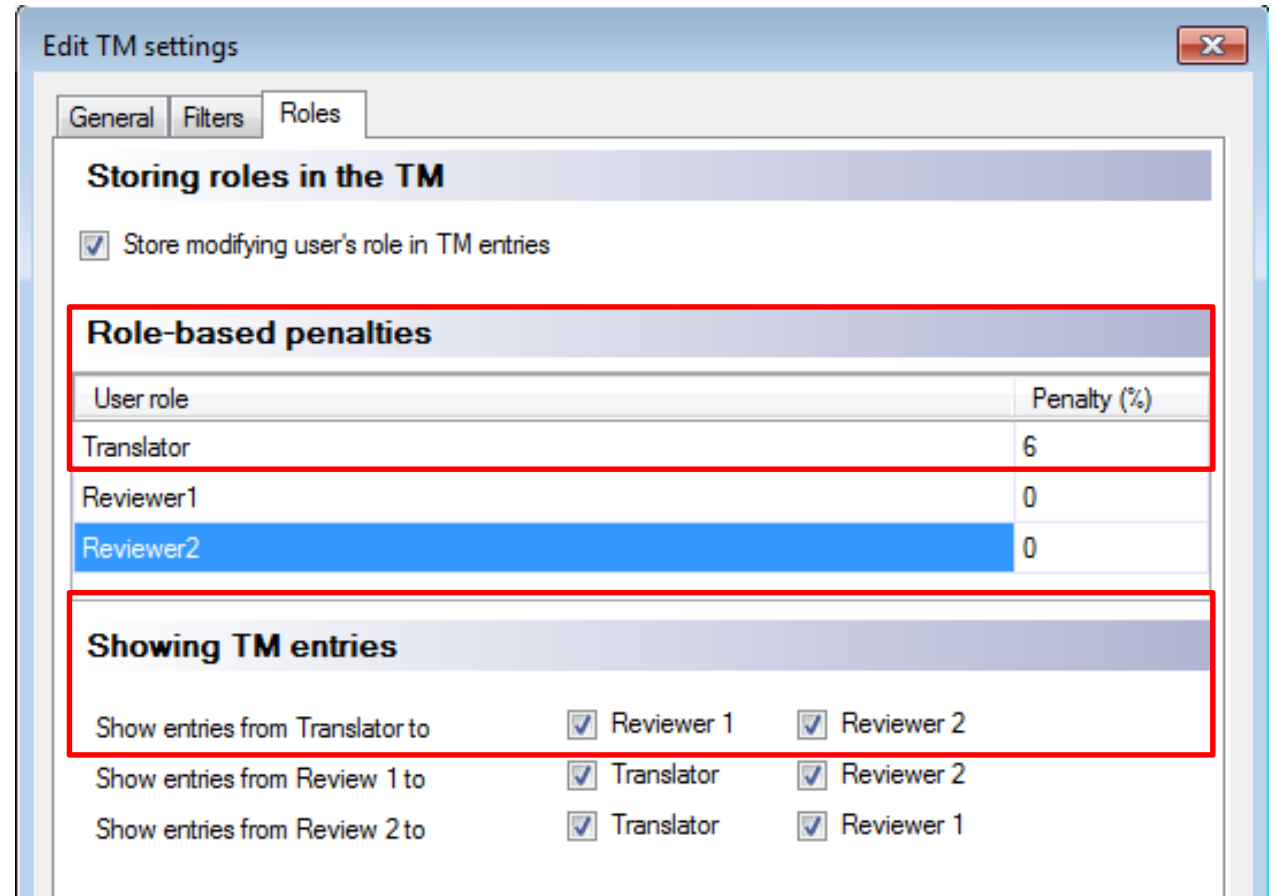○ Applying a penalty to segments with a specific *user name*.

Using Metadata (roles)

○ **Penalty on segments saved with a certain *role*.**

○ **Showing certain matches only to users with a certain *role*.**

Tool-specific data

- Categories like project name, subject matter, client name
- Context of a segment (surrounding text, structural information, associated IDs)
- Document name

Customized data

○ Additional fields in TMs, like

   ○ Content type

      ○ legal, marketing, medical, technical…

      ○ marketing brochures, leaflets, videos, packaging material, banners…

   ○ Document type (website, brochure, manual…)

   ○ Ranking (translated by expert, junior translator…)

   ○ Confirmation status (confirmed by client, confirmed by vendor…)

○ Forbidden (term base)

Using Metadata

○ Use them for filtering, sorting, penalties and…

○ …deciding on the appropriateness of a match
  - ○ Does a translation for marketing fit the context of a contract?
  - ○ If the match was saved to the TM by an internal reviewer, it is probably better to use this one than the one that comes from the alignment done by an intern.

Process Data

○ In the TM
  - ○ Segment in the TM comes from an alignment
  - ○ Segment source has been edited
  - ○ Segment was saved by a user with the role translator, reviewer…
  - ○ Context information
○ In the translation file (XLIFF…)
  - ○ Path to the original file
  - ○ Where the match came from (TM name, MT provider)
  - ○ History of the changes to a segment
  - ○ Comments on a segment
  - ○ Whether segments have been split or merged

Process Data

○ Tool-specific metadata

○ The segment in the TM
comes from an alignment.

Process Data

○ The source segment has been changed.

○ The segment has been confirmed last by a person in the role of reviewer 2.

Segment details	
Created by	angelikazerfass
Created at	25.05.2018 14:32:14
Modified by	angelikazerfass
Modified at	25.05.2018 14:32:14
Last modified role	None
Document	Demo 1.docx
Corrected	Yes
Aligned	No

Segment details	
Created by	angelikazerfass
Created at	25.05.2018 14:35:20
Modified by	angelikazerfass
Modified at	25.05.2018 14:35:33
Last modified role	Reviewer 2
Document	Demo 1.docx
Corrected	No
Aligned	No

Process Data

○ Context information will usually not be usable by other tools, because they use a different way to save context data or save different data.

How tool A saves context information to the TMX

```
<prop type="x-context-pre">previous segment</prop>
<prop type="x-context-post">following segment</prop>
```

How tool B saves context information to the TMX

```
<prop type="x-context">297704054075490337, -2182033961215568804</prop>
<prop type="x-ContextContent">previous segment. |  | previous translation. | </prop>
```

Metadata in Reports

○ Usually, statistics on word counts and match values are created at the start of the project to determine the scope of the project.

○ But there are also reports that associate different kinds of metadata

  ○ Post-translation analysis: What match values (taken from the XLIFF file) did the translator (name) actually receive.

  ○ Edit distance report: Number of changes a user had to make to the translation.

  ○ Change report (reviewer): How many changes did the reviewer make using the tracking changes feature.

  ○ History of a segment or term base entry (who changed the segment in which way when and in what user role).

WHAT ELSE YOUR SYSTEM KNOWS ABOUT YOU

User Names

○ User names are great to determine who saved/confirmed an entry. This can tell the next user whether he/she can trust the suggested match.

○ But as soon as files, TMs, term bases get exchanged with outside users, names might need to be disguised or deleted (also think of GDPR (General Data Protection Regulation)/DSGVO (DE) requirements).

  ○ A vendor who provides TMX or XLIFF files to a client and knows that these will be shared with other vendors does not want the clear name of their translators in the TMX file.

User Name – Computer Name

○ **System data**
(data saved automatically to the TM with every segment)

  ○ User name (how is the user logged into the system)

  ○ Date/time   (time format of your OS)

  ○ Computer name (some systems only)

Ninas_laptop/
flowerpower

Paths

○ Translation tools import files and convert them to their internal bilingual format.

○ After translation the file is exported again.

○ The system keeps the information on the path to the original file.

\<path\>X:\Translation Jobs\Client A\Marketing\2_to translation\Demo1.docx\</path\>

\<value key="OriginalFilePath"\>C:\Users\flowerpower\Desktop\Demo1.docx\</value\>

o Valuable information for troubleshooting if you need to find out where a file comes from or if this file is the same as the one sent out originally.

o But it also shows how the user is organizing the work…

File Names

○ Depending on the options your TM system might save the name of the document with the segments.

`<prop type="filename">nl_Driver.docx</prop>`

`<prop type="filename">test.docx</prop>`

`<prop type="filename">X-Shape_Manual_Author-JoeMiller_Rev-TomOuting_V3.idml</prop>`

Editing Time

○ With more and more post-editing being done, users need to track how much time they spend on a segment.

```
<commitinfo matchrate="80" timestamp="2019-08-30T11:51:06Z" username="user1" editingtime="26599" />
```

○ Although this information is mainly visible for the users themselves it might become visible to project managers in the future as well.

Translation Origin

○ Information on match values, status of the segment or the source of a match.

Summary

- Metadata is everywhere
  - Automatically created, generic and exchangeable
  - Created by the system, but system-specific and therefore not exchangeable
  - Created by the users, partially exchangeable

- The user decides on how the data looks like.
- The user sometimes is not aware how much metadata is saved and in what way.

- Metadata is very useful to
  - decide on the appropriate match
  - troubleshoot
  - view the history of actions

Your opinion is important to us! Please tell us what you thought of the lecture. We look forward to your feedback under

http://lt02.honestly.de

or scan the QR code

Angelika Zerfaß

zerfass@zaac.de

The feedback tool will be available even after the conference!