Raw Machine Translation

A way to make essential information available for potential customers

Sabine Peng
Senior Localization Program Manager
August 2021
Agenda

Why Raw MT
What's the Strategy
How to implement Raw MT
Case Study
Achievements
Recap & What’s Next
Q&A
Why Raw MT?

Cost Efficiency

Technology Readiness

Business Needs
What’s the Strategy?

Part of Tiered Localization
Raw MT is part of Tiered Localization, the solution package with pre-defined and customized solutions for different scenarios.

Docs First
Raw MT is applied to Docs First considering that UI localization is more complex with higher risk.

Business Requirements
Raw MT is requested for some specific Business Requirements.

Specific for Some Locales/Products
Raw MT is used in Selected Languages and Products based on the data analysis.
Tiered Localization Introduction

Business Requirements
Market Data
Production Model
Language Set
Component

Tier-A (MT + PE)
Tier-B (HT)
Tier-C (Raw MT)

BI and Technology (MT, ML, TMS...)

©2021 VMware, Inc.
Business Requirements

Product Documentation
As per customer survey, customers prefer using localized documentation, so Raw MT is implemented in some scenarios.

Hands-on-Lab
Raw MT is required for the updates of localized manuals, quick with no cost.

Knowledge Base
Raw MT is applied to some selected relatively high viewed articles in selected languages, while top viewed articles use MT + PE.

Internal Reference
Raw MT is used for internal document requests.
Selected Languages

Language Candidates

- Brazilian Portuguese
- Italian
- Traditional Chinese
- Danish
- Dutch

Selected Languages

- Traditional Chinese
- Brazilian Portuguese
Selected Products

Top 3 Products Selected
- Product A
- Product B
- Product C
Workflow Readiness

**CMS**
- Create Translation Job
- Generate Output

**Publishing System**
- Publish to VMware Docs with MT Banner

**TMS**
- Protect DNT Terms
  - Feed MT'ed Files to CMS
  - Machine Translate Files
  - Add MT Meta Data
  - Profanity Check
  - MT Engine
  - MT Engine

**CMS**
- Create Translation Job
- Generate Output

**Publishing System**
- Publish to VMware Docs with MT Banner

**TMS**
- Protect DNT Terms
  - Feed MT'ed Files to CMS
  - Machine Translate Files
  - Add MT Meta Data
  - Profanity Check
  - MT Engine
  - MT Engine

**CMS**
- Create Translation Job
- Generate Output

**Publishing System**
- Publish to VMware Docs with MT Banner

**TMS**
- Protect DNT Terms
  - Feed MT'ed Files to CMS
  - Machine Translate Files
  - Add MT Meta Data
  - Profanity Check
  - MT Engine
  - MT Engine

**CMS**
- Create Translation Job
- Generate Output

**Publishing System**
- Publish to VMware Docs with MT Banner

**TMS**
- Protect DNT Terms
  - Feed MT'ed Files to CMS
  - Machine Translate Files
  - Add MT Meta Data
  - Profanity Check
  - MT Engine
  - MT Engine

**CMS**
- Create Translation Job
- Generate Output

**Publishing System**
- Publish to VMware Docs with MT Banner

**TMS**
- Protect DNT Terms
  - Feed MT'ed Files to CMS
  - Machine Translate Files
  - Add MT Meta Data
  - Profanity Check
  - MT Engine
  - MT Engine

**CMS**
- Create Translation Job
- Generate Output

**Publishing System**
- Publish to VMware Docs with MT Banner

**TMS**
- Protect DNT Terms
  - Feed MT'ed Files to CMS
  - Machine Translate Files
  - Add MT Meta Data
  - Profanity Check
  - MT Engine
  - MT Engine
Raw MT Output Evaluation

Prerequisite - Prediction Model in Place
- Golden translation
- Normal translation – PED% ≤ 20%
- Bad translation – PED% > 20%

Formula:
- Overall score = Sum of Prediction Score for each Segment ÷ No. of Segments
- Validation Result - Pass

<table>
<thead>
<tr>
<th>Project Name</th>
<th>Locale</th>
<th>Workgroup</th>
<th>Project #</th>
<th>Creation Date</th>
<th>Description</th>
<th>Overall Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>test score3333</td>
<td>de_DE_utf16</td>
<td>R&amp;D G11n</td>
<td>72293</td>
<td>11/11/20, 10:30 PM</td>
<td></td>
<td>80</td>
</tr>
<tr>
<td>test score222</td>
<td>fr_FR_utf16</td>
<td>R&amp;D G11n</td>
<td>72282</td>
<td>11/11/20, 10:26 PM</td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>test score</td>
<td>pl_PL_utf16</td>
<td>R&amp;D G11n</td>
<td>72281</td>
<td>11/11/20, 10:22 PM</td>
<td></td>
<td>50</td>
</tr>
</tbody>
</table>
Sample
Case Study 1 – CN2TW

TW Page View Trend
The trend of TW page view is increasing after Raw MT roll-out.

Page View Growth Rate Trend
TW page view growth rate is almost equal to or higher than EN page view growth rate after Raw MT roll-out.

Q/Q Growth rate = (page view this month – page view last month) / page view last month
Case Study 2 – EN2BR

**BR Page View Trend**
The trend of BR page view is increasing after Raw MT roll-out.

**Page View Growth Rate Trend**
BR page view growth rate is higher than EN page view growth rate most of the time after Raw MT roll-out.

Q/Q Growth rate = (page view this month – page view last month) / page view last month
Achievements

BU Quotes
“Thanks for the update! Good to know this project is being used by our users.”
“I’m looking forward to the page view data of MT translation 😊”

Cost Avoidance
Cost avoidance takes -7% of the total cost last fiscal year.

Page View Growth
Page view increases over 100% after Raw MT roll-out.
Recap & What’s Next

2019
- Analyzed data
- Set up workflow
- Started pilot for CN2TW

2020
- Started new language PT-BR
- Optimized MT quality
- Rolled out Raw MT for EN2BR

2021
- Deploy customer rating feature
- Continue with quality optimization
- Roll out APE aided Raw MT

2022
- Smart MT (Raw MT or MT PE, decided by ML)
- Explore more Raw MT use cases
Thank You