



Improving quality of Machine Translation output for tag heavy text



 \rightarrow

JMF



xi:include href='sas.xis.doc.xml://commonbivwr/xml/common_element_includes.xml' xmlns:xi=' xi:fallback xi:include The required URI is softwareFeature /folders/folders/@myFavorites softwareFeature .
xi:include href='sas.xis.doc.xml://commonbivwr/xml/common_element_includes.xml' xmlns:xi=' xi:fallback xi:include The required capability is softwareFeature Personalization softwareFeature .
listUnordered xi:include href='sas.xis.doc.xml://commonbivwr/xml/common_element_includes.xml' xmlns:xi=' xi:fallback xi:include
A Report Tile Does Not Display an Image
Problem:
When you are in windowiten conditions="ins I win" My Reports windowiten

memoQ Interna



key The creative **COP** cannot be added because it contains images that are encoded in Base64, which is not supported.

For more information about the diagnostics agent, see the <a href=\"https://documentation.sas.com/?softwareId=ONEMKTMID&softwareVersion=produ Diagnostics Access Points topic in the Help Center.

For more information about the direct marketing agent, see the <a href=\"https://documentation.sas.com/?softwareId=ONEMKTMID&softwareVersion=produ Direct Access Points topic in the Help Center.





Improving quality of Machine Translation output for tag heavy text



 \rightarrow

JMF





Problems with MT



1

Within sentence tags in source text are a problem for MT.

Quality of MT output

2

For very domain specific source text, off the shelf, generic MT models provide inaccurate translations.

Validation and postediting

3

QA and postediting of MT output is high cost.



Tags in source and MT results

Use both formatting and tags VS. Use text and formatting

Formatting and lookup

Use text and formatting

Use plain text only

Use both formattings and tags

Use both formattings and tags

	Google Cloud Translation Advanced plugin settings					
e memoqfest	Authentication Select the service acco	unt's .json file you downloaded from Google Cloud Platforr	n.			
MT results	Browse	C:\temp\MT\jmp-nmt-780f63c1fff6.json				
	Regional endpoint	us-central1	\sim			
g and tags	AutoML custom mod Custom models: dataset_167144892_2	dels and glossary 0221219115733	~			
	Choose one MT mod custom model for a la	el per language pair. If you don't choose, or there is no anguage pair, Google will use the general model.				
matting	Glossary: No glossary available					
	If you can't find your	glossary in the list, create one.				
	-Formatting and look	up				
and lookup						
ormattings and tags			\sim			
ext only						
d formatting						
ormattings and tags						
slate segments that match th	ne tollowing	redex:	_			

redex:

OK	Cancel	H

elp





Tags in source and MT results

62	Click on windowItem Refine Task Type windowItem and select the windowItem Create a recommendation task that delivers personalized content windowItem option.	Pulse en windowItem Refinar el tipo de tarea windowItem y seleccione el windowItem Crear una tarea de recomendación que entregue contenido personalizado windowItem Opción	0%	×
63	Click on "Refine Task Type" and select the "Create a recommendation task that delivers personalized content" option.	Haga clic en "Refinar tipo de tarea" y seleccione la opción "Crear una tarea de recomendación que entregue contenido personalizado".	0%	×
	Clicking windowItem Create Task windowItem before setting the	Haciendo clic windowItem Crear tarea windowItem antes de establecer las	0%	×
64	recommendations options in windowItem Refine Task Type windowItem creates a standard task.	opciones de recomendaciones en windowItem > Refinar el tipo de tarea		2
65	Clicking "Create Task" before setting the recommendations options in "Refine Task Type" creates a standard task.	Al hacer clic en "Crear tarea" antes de establecer las opciones de recomendaciones en "Refinar tipo de tarea" se crea una tarea estándar.	0%	×
66	After clicking windowItem Create Task windowItem , you cannot return to this screen to select the option to create a recommendation task.	Después de hacer clic windowItem Crear tarea windowItem , no puede volver a esta pantalla para seleccionar la opción para crear una tarea de recomendación.	0%	×
67	After clicking "Create Task", you cannot return to this screen to select the option to create a recommendation task.	Después de hacer clic en "Crear tarea", no puede volver a esta pantalla para seleccionar la opción para crear una tarea de recomendación.	0%	×





Quality and accuracy of MT

English

Once your web task is set up, you must change the status in the "Orchestration" tab to "Mark Ready" or "Mark Ready and Publish" before your task can become active.

Spanish - Custom model

Una vez configurada la tarea web, debe cambiar el estado en la pestaña "Orquestación" a "Marcar como lista" o "Marcar como lista y publicar" para que la tarea se pueda activar.

Spanish - Google NMT model

Una vez que su tarea web esté configurada, debe cambiar el estado en la pestaña "Orquestación" a "Marcar listo" o "Marcar listo y publicar" antes de que su tarea pueda activarse.

TRANSLATE





Solution

Normalize tags

1

Replace tags with non-semantic elements in order not to break the sentence structure and produce grammatically correct translations.

Train custom NMT model

2

Aggregate and cleanse TM assets to train custom, domain specific NMT models.

Enhanced QA

3

Use replacements and RegEx checks to reduce postediting efforts.



Source hash file





<item eid="p0531f2lhroqaon18c9nlpnrf8bm"><paragraph eid=
"n14tqgvctxi3wun15rhixno4rizs">Click the <windowItem>Overview</windowItem>
tab to see how key metrics performed by total users in the task.
</paragraph>CRUE

<paragraph eid="n070elimugga22n13nuqe19c3rcy">Click <inlineGraphic><graphicAndDescription><altImageDescription>edit </altImageDescription><graphic scalefit="0" scalefitwidth="9.9pt" source="sas.xis.graphic://commoniconshtml/image/edit.svg"/> </graphicAndDescription></inlineGraphic> for the <windowItem>Date </windowItem> field to specify the range of dates for the data that you want to see. You can select dates from the start date to the end date for the data that is available for the test task. For more information about using the report and the data in the report, click <windowItem>Details</windowItem>.</paragraph><listUnordered eid= "n11d1gfjkxcwk6n1jtb882fr8ylp">CRIM

<item eid="n0yuk6f9zgci12n1dqrxmchxabjn">CRLF





```
#19@:<inlineGraphic>
#20@:<graphicAndDescription>
#21@:<graphic scalefit="0" scalefitwidth="9.9pt"
source="sas.xis.graphic://commoniconshtml/image/edit.svg"/>
#22@:</graphicAndDescription>
#23@:</inlineGraphic>
#24@:<listUnordered eid="nlldlgfjkxcwk6nljtb882fr8ylp">
```





Reference tags

\$line =~ s/<windowItem>/+ /g;
\$line =~ s/<\/windowItem>/ #/g;
\$line =~ s/<altImage>/\$ /g;
\$line =~ s/<\/altImage>/ ¥/g;

<windowName>Regression analysis</windowName>
->
+ Regression analysis #

Formatting and other non-reference tags \$line =~ s/\Q\$_\E/\#\$count\@/;

2

Click<inGraphic><inDesc><altImage>Select Event</altImage> -> Click #96@#97@¢ Select Event ¥#98@



Normalize tags

1

.hsh files

3



#26@:<listUnordered eid="p1cl2xhlo12dkqn1dy6a4m8pkvdy">
#27@:<xrefSee link="p0gr5dhxsmzduon1qovu5u440dvj">
#28@:<overrideFetchedText>
#29@:</overrideFetchedText>
#30@:<xrefText/>
#31@:</xrefSee>
#32@:<xrefSee link="n0hx4toju8d2vyn1okwepaok3ucc"
targetComment="Chapter 30, "Types of Data Available for
Targeting,"" targetFile="cintug/xml/chapters/ch-targeting-data.xml">





Pre-processed source with normalized tags



59.	Click [†] Refine Task Type [‡] and select the [†] Create a recommendation task that delivers personalized content [‡] option.	Haga clic en † Refinar tipo de tarea ‡ y seleccione la opción † Crear una tarea de recomendación que entregue contenido personalizado ‡.	0%	×
60.	Clicking [†] Create Task [‡] before setting the recommendations options in [†] Refine Task Type [‡] creates a standard task.	Al hacer clic en † Crear tarea ‡ antes de establecer las opciones de recomendaciones en † Refinar tipo de tarea ‡ se crea una tarea estándar.	0%	×
61.	After clicking [†] Create Task [‡] , you cannot return to this screen to select the option to create a recommendation task.	Después de hacer clic en † Crear tarea ‡, no puede regresar a esta pantalla para seleccionar la opción para crear una tarea de recomendación.	0%	×

55.	#55@: <xrefsee cintug="" link="n028xtjvaysj35n116gtas14gvqu" personalization="" prsnlzn-add-merge-tag.xml"="" targetcomment="
Personalized Information to Creatives
targetFile=" xml=""></xrefsee>	Adding	#55@: <xrefsee <br="" link="n028xtjvaysj35n116gtas14gvqu" targetcomment="Añadir
información personalizada a creatividades ">targetFile="cintug/xml/personalization/prsnlzn-add-merge-tag.xml"></xrefsee>	0%	×
				Changed n	ever
View	pane			1	」 中

View pane

#54@:<listUnordered eid="n1c9x9fpmzd2tpn1xbizy18o32vg">

#55@:<xrefSee link="n028xtjvaysj35n116gtas14gvqu" targetComment="Adding Personalized Information to Creatives" targetFile="cintug/xml/personalization/prsnlzn-add-merge-tag.xml">





Training custom MT models

≡ Google Cloud 🔹 sas	AutoML 🔻		Q Search Products, resources, docs (/)		. 0	: 📵							
ネA Translation	← dataset_16470021474	477	II, VIEW STATS 🏦 EXPORT DATA										
Dashboard	IMPORT SENTENCES	IMPORT SENTENCES TRAIN EVALUATE PREDICT											
Datasets	Import completed with partial	errors.											
P Models	Enter property name or value												
	in denerous.	024,000	Source	Target	Set								
	Training	507,722	Avail. Credit:	Credito disp.:	Train	iing							
			Source Table Name	Nome della tabella di origine	Train	ing							
	Validation	8,502	You must select either users or groups.	Selezionare gli utenti o i gruppi.	Train	ning							
	Testing	8,426	specifies the host on which the application server that is used to connect to SAP resides.	specifica l'host su cui risiede l'application server che viene utilizzato per connettersi a SAP.	Train	ling							
			should be a positive integer.	dovrebbe essere un numero intero positivo.	Train	iing							
	Tilter Filter file		No Redshift metrics were retrieved from cloud watch.	Nessuna metrica Redshift è stata recuperata da cloud watch.	Train	ning							
Validation		Model names not listed in inventory Nomi modelli non elencati nel catalogo				ing							
-	en_it_val_tsv	8,502	The Add-in for Microsoft Office was unable to locate a connection to a Web Report Studio (WRS) web service from this repository, which is necessary to open a report. Either this feature was not turned on in WRS, or the web service is not running. Please contact your System Administrator.	L'Add-in for Microsoft Office non è stato in grado di individuare una connessione a un servizio Web di Web Report Studio (WRS) da questo repository, che è necessario per aprire un report. Questa funzionalità non è stata attivata in WRS o il servizio Web non è in esecuzione. Contattare l'amministratore del sistema.	Train	ing							
	Testing		Column mapping:	Mappatura delle colonne:	Train	ing							
		Il sistema operativo minimo supportato è iOS 14.5 o versione successiva.	Train	iing									
-	en_it_test_tsv	Il fattore di ponderazione del rischio per esposizioni cartolarizzate relative a portafogli non diversificati nella 6° categoria RW (%):	Train	ling									
	Training		Perform secondary objective range analysis:	Esegui analisi del range dell'obiettivo secondario:	Train	ing							
			Data Set Labels	Etichette dei data set	Train	ing							
-	en_it_tsv	507,722	Removing Items	Rimozione di elementi	Train	iing							
		To sort reports, tap Sort button and make a selection. Per ordinare i report, tocca Pulsante Ordina ed effettua una selezione.				ing							
			the command to start a SAS session on the server	il comando per avviare una sessione SAS sul server	Train	iing							
			In-Control Average Run Length	Lunghezza esecuzione media sotto controllo	Train	ling							





Custom model vs. "Raw" MT



BLEU Score gains



Raw Google Custom model



End game



Removing extra spaces

Find and replace extra spaces around replacement tag elements (e.g., # 67 @ -> #67@)



QA and review in memoQ

Counting tag replacements Terminology check against UI



Post-processing

Resolving tag replacements Putting tags from hsh files into output xml files





Enhanced QA

Edit QA settings

Segments and terms	Consistency	Numbers	Punctuation	Spaces, capita	ls, characters	Inline tags	Length	Regex	Severity	
🗹 Use regular expre	ession checks									
Wam if	missing rege	ex replacem	ent in target			\sim				
Source regex	(\#[0-9]+	·\@)							Rx	
Replacement	\$1								Rx	
Correction									Rx	
Description	Missing has	h in target								
	Expand t	tags to text	before process	sing						
	Add		Update	Delete	Clear all					
 Image: Missing tage Image: Missing tage Image: Missing tage Image: Source Target Image: Tags: - co Image: Tags: - co<td>gs" - missing re ash in target" - r regex: (\#[0-9] replacement: \$ ame" - counts of ounts of regex n counts of regex ' - forbidden reg " - forbidden reg " - forbidden reg " - forbidden reg m" - counts of r m" - counts of r</td><td>gex replace missing rege [+\@) if regex matches diff x matches diff x matches diff x matches diff x matches diff x matches diff counts of reger regex match regex match regex match</td><td>ement in target ex replacement ches differ er differ n target in target gex matches differ ex matches differ</td><td>t in target liffer fer</td><td></td><td></td><td></td><td></td><td></td><td></td>	gs" - missing re ash in target" - r regex: (\#[0-9] replacement: \$ ame" - counts of ounts of regex n counts of regex ' - forbidden reg " - forbidden reg " - forbidden reg " - forbidden reg m" - counts of r m" - counts of r	gex replace missing rege [+\@) if regex matches diff x matches diff x matches diff x matches diff x matches diff x matches diff counts of reger regex match regex match regex match	ement in target ex replacement ches differ er differ n target in target gex matches differ ex matches differ	t in target liffer fer						
					OK	(Cancel		Help	

Х



Enhanced QA

Edit QA settings										×	
Segments and terms	Consistency	Numbers	Punctuation	Spaces, capit	als, characters	Inline tags	Length	Regex	Severity		
Use regular expression checks											
Wam if	counts of re	counts of regex matches differ \checkmark									
Source regex	(†){1,}								Rx		
Target regex	(†){1,}								Rx		
Correction									Rx		
Description	windowltem										
	Expand	tags to text	before process	sing							
	Add		Update	Delete	Clear all						
Source	regex: (\#[0-9]+\@)							,	^	
	ame" - counts o	of regex mat	ches differ						- 1		
Tags" - co	ounts of regex r	matches diff	er								
	-forbidden rej	gex match i	n target								
🕀 🧖 "Missing @	" - forbidden re	egex match	in target								
🗄 🦓 "Closing wi	ndowName" - (counts of re	gex matches d	liffer							
Source	recev: (t){1}	regex matc	hes differ								
Target	regex: (t){1,}										
🗄 🥋 "Closing wi	ndowltem" - co	ounts of reg	ex matches diff	fer							
					OF	(Cancel		Help		









Beautify and normalize tag heavy source









Cleanup and maintain healthy TMs and TBs







Train your own NMT models

memoQ Interna









Develop advanced QA profiles for normalized rags

.







Thank you!

Any questions?

Krzysztof.Jozefowicz@jmp.com