D3.2.1: OKAPI OCELOT

Phil Ritchie

Distribution: Consortium Internal Report

MultilingualWeb-LT (LT-Web)
Language Technology in the Web

FP7-ICT-2011-7

Project no: 287815

Document Information

Deliverable number:	3.2.1
Deliverable title:	Report on VistaTEC deliverables
Dissemination level:	CO
Contractual date of delivery:	31st March 2012
Actual date of delivery:	08/12/2013
Author(s):	Phil Ritchie, CTO, VistaTEC
Participants:	UL, DFKI
Internal Reviewer:	TCD
Workpackage:	WP1
Task Responsible:	@@@@
Workpackage Leader:	Felix Sasaki

Revision History

Revision	Date	Author	Organization	Description
1	11/09/2013	Phil Ritchie	VistaTEC	Initial version.
2	08/12/2013	Phil Ritchie	VistaTEC	Final version.

Contents

1. Executive Summary

VistaTEC's primary output from this project is a platform independent desktop editor called "Ocelot". The editor can read/write xliff+its files; is based on the Okapi framework and thus integrates seamlessly with the Okapi workflow.

Ocelot implements the Language Quality, MT Confidence and Provenance data categories of ITS 2.0. Metadata from these categories that is attached to translation units in the XLIFF file can be rendered alongside those segments in a user-definable way through a mechanism of rendering rules. In addition to rendering the metadata, the rules can be used to filter segments according to values of the metadata.

The benefit of a translator or post-editor being able to see this metadata within their editing environment is that their task can be directed and made more efficient: hints about errors and information of a segment's provenance can aid the translator in deciding how to edit a segment, if at all.

Ocelot is written in Java which makes it platform independent allowing for maximum deployment. See *"Ocelot Open Source And Technical Details"*.

2. Ocelot Features

Ocelot has the following features:

- Reads/writes Language Quality Issue and Provenance ITS 2.0 data categories. Also reads MT Confidence data category.
- User configurable rules for defining how metadata is rendered in the sidebar alongside each segment.
- User configurable rules for defining how segments can be filtered according to properties of the metadata.
- User configurable keyboard shortcuts for adding new Language Quality Issue metadata.

3. Ocelot Benefits

Ocelot increases the efficiency and accuracy of the linguistic review and post-editing processes by providing translators with access to useful contextual metadata that they may not otherwise have. Metadata gathered at earlier stages of the localization process, such as, automated translation; automated quality assurance checks; NLP-based analytical processes; etc. can inform and direct the translator as to actions that they might want to take.

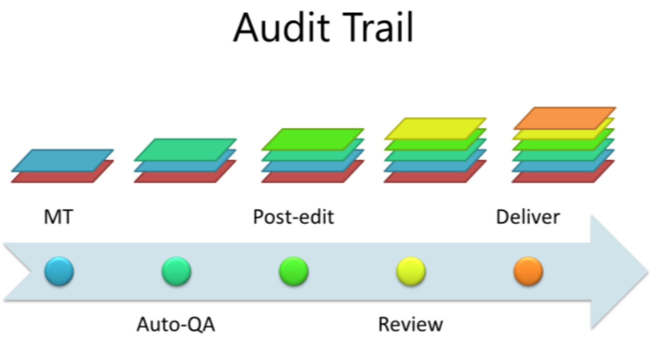
By way of a concrete example:

A document may be translated partially using machine translation and human. At this stage of the process the following Provenance metadata could be gathered:

- Machine translation output confidence scores,
- Name and version of the machine translation engine,
- Name of the human translator and the organization they work for.

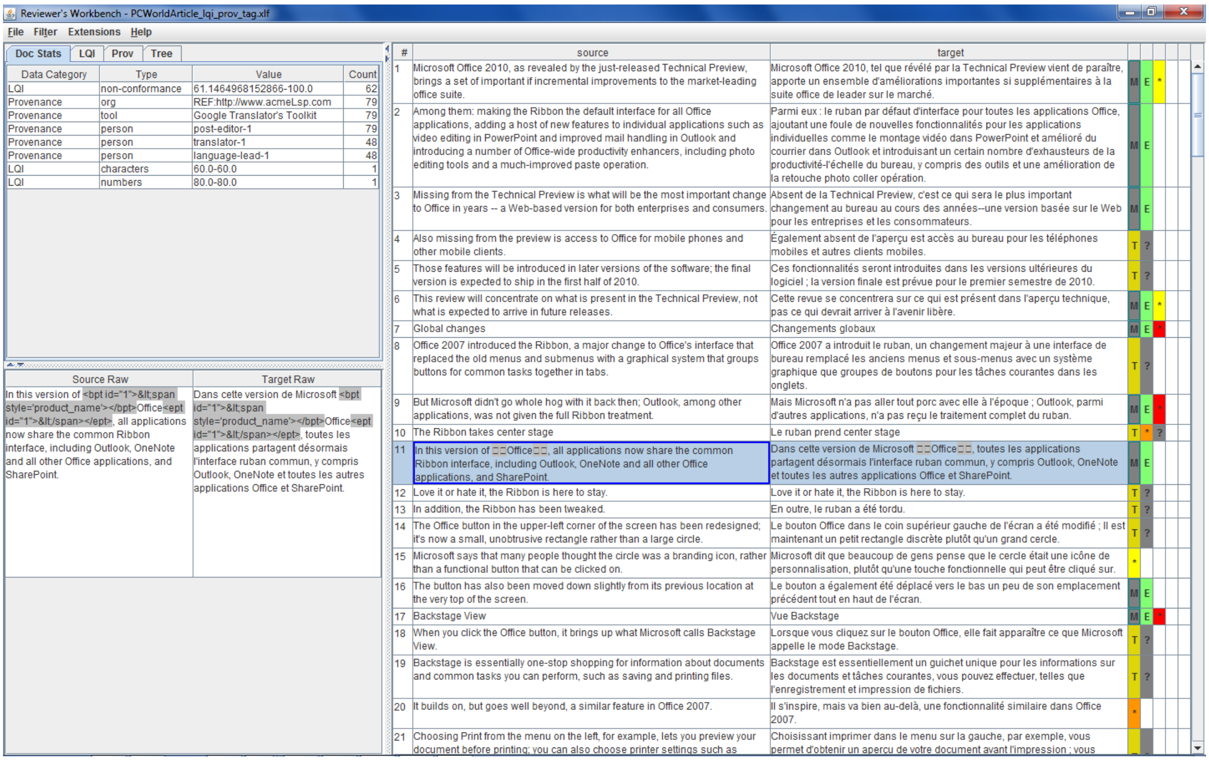
Following translation the document could be passed through an automated quality assurance pipeline. During this phase of the localization process the metadata listed below could be added to the segments:

- Name and version of the automated QA programs,
- Types and severity of errors found,
- NLP/Text Classification hypotheses as to the quality and suitability of the translations.



Thus when the document arrives at the desk of a linguistic reviewer, all of this information is at their finger tips. This can enable them to make decisions as to the strategy of their task: address errors first according to severity level, review only those segments proposed by machine translation, specifically review segments proposed by a translator who is known to be a novice, etc.

4. Ocelot User Interface

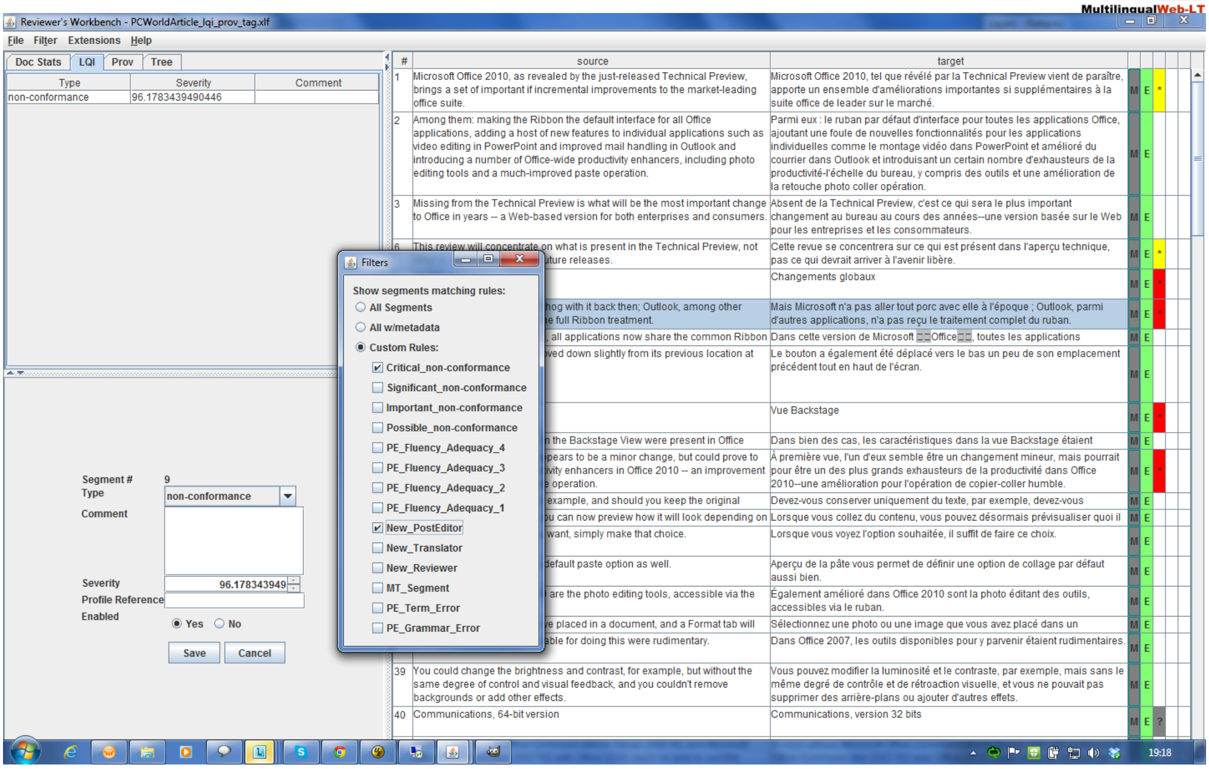


Data Category	Type	Value	Count
LQI	non-conformance	61.1464968152866-100.0	62
Provenance	org	REF:http://www.acmeLsp.com	79
Provenance	tool	Google Translator's Toolkit	79
Provenance	person	post-editor-1	79
Provenance	person	translator-1	48
Provenance	person	language-lead-1	48
LQI	characters	60.0-60.0	1
LQI	numbers	80.0-80.0	1

Source Raw	Target Raw
In this version of <bpt id="1">&lt;span style='product_name'></bpt>Office<ept id="1">&lt;/span></ept>, all applications now share the common Ribbon interface, including Outlook, OneNote and all other Office applications, and SharePoint.	Dans cette version de Microsoft <bpt id="1">&lt;span style='product_name'></bpt>Office<ept id="1">&lt;/span></ept>, toutes les applications partagent désormais l'interface ruban commun, y compris Outlook, OneNote et toutes les autres applications Office et SharePoint.

#	source	target			
1	Microsoft Office 2010, as revealed by the just-released Technical Preview, brings a set of important if incremental improvements to the market-leading office suite.	Microsoft Office 2010, tel que révélé par la Technical Preview vient de paraître, apporte un ensemble d'améliorations importantes si supplémentaires à la suite office de leader sur le marché.	M	E	*
2	Among them: making the Ribbon the default interface for all Office applications, adding a host of new features to individual applications such as video editing in PowerPoint and improved mail handling in Outlook and introducing a number of Office-wide productivity enhancers, including photo editing tools and a much-improved paste operation.	Parmi eux : le ruban par défaut d'interface pour toutes les applications Office, ajoutant une foule de nouvelles fonctionnalités pour les applications individuelles comme le montage vidéo dans PowerPoint et amélioré du courrier dans Outlook et introduisant un certain nombre d'exhausteurs de la productivité-l'échelle du bureau, y compris des outils et une amélioration de la retouche photo coller opération.	M	E	
3	Missing from the Technical Preview is what will be the most important change to Office in years -- a Web-based version for both enterprises and consumers.	Absent de la Technical Preview, c'est ce qui sera le plus important changement au bureau au cours des années--une version basée sur le Web pour les entreprises et les consommateurs.	M	E	
4	Also missing from the preview is access to Office for mobile phones and other mobile clients.	Également absent de l'aperçu est accès au bureau pour les téléphones mobiles et autres clients mobiles.	T	?	
5	Those features will be introduced in later versions of the software; the final version is expected to ship in the first half of 2010.	Ces fonctionnalités seront introduites dans les versions ultérieures du logiciel ; la version finale est prévue pour le premier semestre de 2010.	T	?	
6	This review will concentrate on what is present in the Technical Preview, not what is expected to arrive in future releases.	Cette revue se concentrera sur ce qui est présent dans l'aperçu technique, pas ce qui devrait arriver à l'avenir libère.	M	E	*
7	Global changes	Changements globaux	M	E	*
8	Office 2007 introduced the Ribbon, a major change to Office's interface that replaced the old menus and submenus with a graphical system that groups buttons for common tasks together in tabs.	Office 2007 a introduit le ruban, un changement majeur à une interface de bureau remplacé les anciens menus et sous-menus avec un système graphique que groupes de boutons pour les tâches courantes dans les onglets.	T	?	
9	But Microsoft didn't go whole hog with it back then; Outlook, among other applications, was not given the full Ribbon treatment.	Mais Microsoft n'a pas aller tout porc avec elle à l'époque ; Outlook, parmi d'autres applications, n'a pas reçu le traitement complet du ruban.	M	E	*
10	The Ribbon takes center stage	Le ruban prend center stage	T	*	?
11	In this version of □□Office□□, all applications now share the common Ribbon interface, including Outlook, OneNote and all other Office applications, and SharePoint.	Dans cette version de Microsoft □□Office□□, toutes les applications partagent désormais l'interface ruban commun, y compris Outlook, OneNote et toutes les autres applications Office et SharePoint.	M	E	
12	Love it or hate it, the Ribbon is here to stay.	Love it or hate it, the Ribbon is here to stay.	T	?	
13	In addition, the Ribbon has been tweaked.	En outre, le ruban a été tordu.	T	?	
14	The Office button in the upper-left corner of the screen has been redesigned; it's now a small, unobtrusive rectangle rather than a large circle.	Le bouton Office dans le coin supérieur gauche de l'écran a été modifié ; Il est maintenant un petit rectangle discrète plutôt qu'un grand cercle.	T	?	
15	Microsoft says that many people thought the circle was a branding icon, rather than a functional button that can be clicked on.	Microsoft dit que beaucoup de gens pense que le cercle était une icône de personnalisation, plutôt qu'une touche fonctionnelle qui peut être cliqué sur.	*		
16	The button has also been moved down slightly from its previous location at the very top of the screen.	Le bouton a également été déplacé vers le bas un peu de son emplacement précédent tout en haut de l'écran.	M	E	
17	Backstage View	Vue Backstage	M	E	*
18	When you click the Office button, it brings up what Microsoft calls Backstage View.	Lorsque vous cliquez sur le bouton Office, elle fait apparaître ce que Microsoft appelle le mode Backstage.	T	?	
19	Backstage is essentially one-stop shopping for information about documents and common tasks you can perform, such as saving and printing files.	Backstage est essentiellement un guichet unique pour les informations sur les documents et tâches courantes, vous pouvez effectuer, telles que l'enregistrement et impression de fichiers.	T	?	
20	It builds on, but goes well beyond, a similar feature in Office 2007.	Il s'inspire, mais va bien au-delà, une fonctionnalité similaire dans Office 2007.	*		
21	Choosing Print from the menu on the left, for example, lets you preview your document before printing; you can also choose printer settings such as	Choisissant imprimer dans le menu sur la gauche, par exemple, vous permet d'obtenir un aperçu de votre document avant l'impression ; vous			

Figure 1: Main Workbench window. Top left-hand panel shows summary of all metadata found in the file. Bottom left-hand panel shows segment with raw mark-up. Left-hand side of main window shows metadata rendered according to user-defined rules.



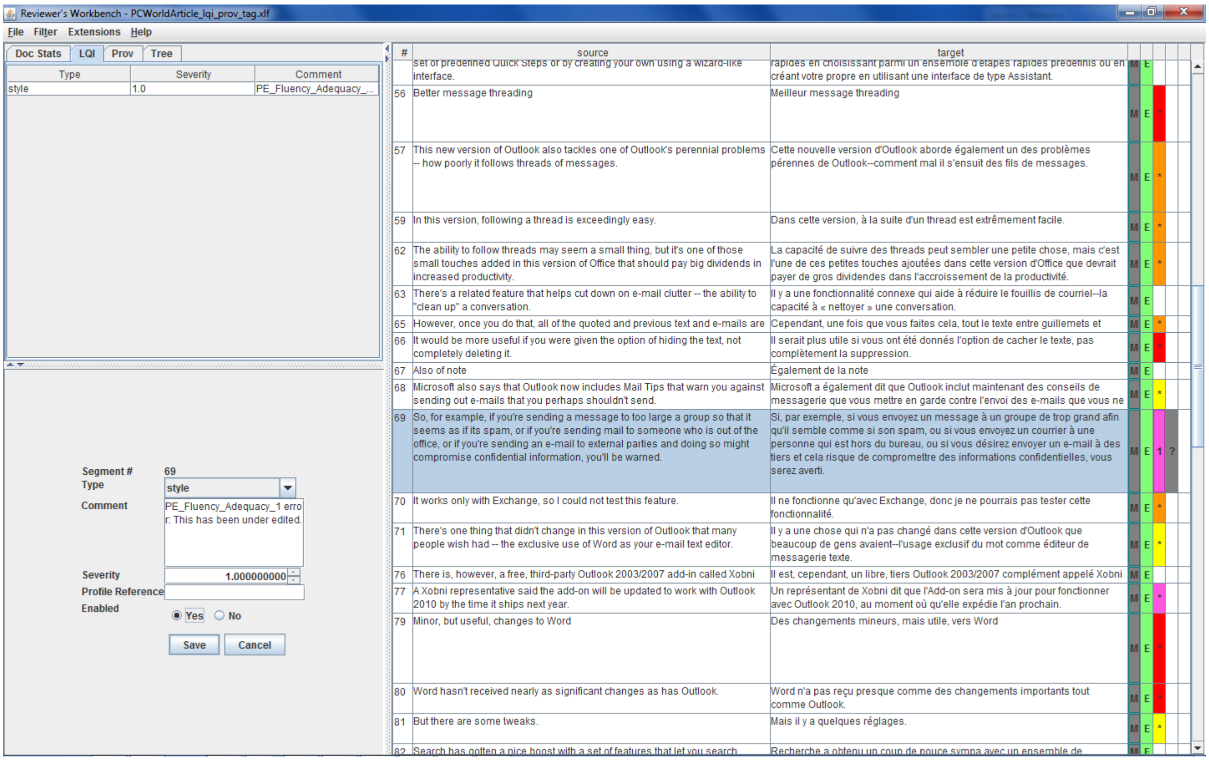
Type	Severity	Comment
non-conformance	96.1783439490446	

Segment # 9
Type non-conformance
Comment
Severity 96.178343949
Profile Reference
Enabled ● Yes ○ No

Show segments matching rules:
○ All Segments
○ All w/metadata
● Custom Rules:
☑ Critical_non-conformance
☐ Significant_non-conformance
☐ Important_non-conformance
☐ Possible_non-conformance
☐ PE_Fluency_Adequacy_4
☐ PE_Fluency_Adequacy_3
☐ PE_Fluency_Adequacy_2
☐ PE_Fluency_Adequacy_1
☑ New_PostEditor
☐ New_Translator
☐ New_Reviewer
☐ MT_Segment
☐ PE_Term_Error
☐ PE_Grammar_Error

#	source	target			
1	Microsoft Office 2010, as revealed by the just-released Technical Preview, brings a set of important if incremental improvements to the market-leading office suite.	Microsoft Office 2010, tel que révélé par la Technical Preview vient de paraître, apporte un ensemble d'améliorations importantes si supplémentaires à la suite office de leader sur le marché.	M	E	*
2	Among them: making the Ribbon the default interface for all Office applications, adding a host of new features to individual applications such as video editing in PowerPoint and improved mail handling in Outlook and introducing a number of Office-wide productivity enhancers, including photo editing tools and a much-improved paste operation.	Parmi eux : le ruban par défaut d'interface pour toutes les applications Office, ajoutant une foule de nouvelles fonctionnalités pour les applications individuelles comme le montage vidéo dans PowerPoint et amélioré du courrier dans Outlook et introduisant un certain nombre d'exhausteurs de la productivité-l'échelle du bureau, y compris des outils et une amélioration de la retouche photo coller opération.	M	E	
3	Missing from the Technical Preview is what will be the most important change to Office in years -- a Web-based version for both enterprises and consumers.	Absent de la Technical Preview, c'est ce qui sera le plus important changement au bureau au cours des années--une version basée sur le Web pour les entreprises et les consommateurs.	M	E	
6	This review will concentrate on what is present in the Technical Preview, not ...ture releases.	Cette revue se concentrera sur ce qui est présent dans l'aperçu technique, pas ce qui devrait arriver à l'avenir libère.	M	E	*
		Changements globaux	M	E	*
	...hog with it back then; Outlook, among other ...e full Ribbon treatment.	Mais Microsoft n'a pas aller tout porc avec elle à l'époque ; Outlook, parmi d'autres applications, n'a pas reçu le traitement complet du ruban.	M	E	*
	, all applications now share the common Ribbon	Dans cette version de Microsoft □□Office□□, toutes les applications	M	E	
	...ved down slightly from its previous location at	Le bouton a également été déplacé vers le bas un peu de son emplacement précédent tout en haut de l'écran.	M	E	
		Vue Backstage	M	E	*
	...n the Backstage View were present in Office	Dans bien des cas, les caractéristiques dans la vue Backstage étaient	M	E	
	...pears to be a minor change, but could prove to ...ivity enhancers in Office 2010 -- an improvement ...e operation.	À première vue, l'un d'eux semble être un changement mineur, mais pourrait pour être un des plus grands exhausteurs de la productivité dans Office 2010--une amélioration pour l'opération de copier-coller humble.	M	E	*
	...example, and should you keep the original	Devez-vous conserver uniquement du texte, par exemple, devez-vous	M	E	
	...u can now preview how it will look depending on	Lorsque vous collez du contenu, vous pouvez désormais prévisualiser quoi il	M	E	
	...want, simply make that choice.	Lorsque vous voyez l'option souhaitée, il suffit de faire ce choix.	M	E	
	...default paste option as well.	Aperçu de la pâte vous permet de définir une option de collage par défaut aussi bien.	M	E	
	...are the photo editing tools, accessible via the	Également amélioré dans Office 2010 sont la photo éditant des outils, accessibles via le ruban.	M	E	
	...e placed in a document, and a Format tab will	Sélectionnez une photo ou une image que vous avez placé dans un	M	E	
	...able for doing this were rudimentary.	Dans Office 2007, les outils disponibles pour y parvenir étaient rudimentaires.	M	E	
39	You could change the brightness and contrast, for example, but without the same degree of control and visual feedback, and you couldn't remove backgrounds or add other effects.	Vous pouvez modifier la luminosité et le contraste, par exemple, mais sans le même degré de contrôle et de rétroaction visuelle, et vous ne pouvait pas supprimer des arrière-plans ou ajouter d'autres effets.	M	E	
40	Communications, 64-bit version	Communications, version 32 bits	M	E	?

Figure 2: Popup window shows rules associated with particular metadata characteristics used as filtering rules.



Type	Severity	Comment
style	1.0	PE_Fluency_Adequacy_...

Segment # 69
Type style
Comment PE_Fluency_Adequacy_1 error: This has been under edited.
Severity 1.000000000
Profile Reference
Enabled ● Yes ○ No

#	source	target				
	set of predefined Quick Steps or by creating your own using a wizard-like interface.	rapides en choisissant parmi un ensemble d'étapes rapides prédéfinis ou en créant votre propre en utilisant une interface de type Assistant.	M	E		
56	Better message threading	Meilleur message threading	M	E	*	
57	This new version of Outlook also tackles one of Outlook's perennial problems -- how poorly it follows threads of messages.	Cette nouvelle version d'Outlook aborde également un des problèmes pérennes de Outlook--comment mal il s'ensuit des fils de messages.	M	E	*	
59	In this version, following a thread is exceedingly easy.	Dans cette version, à la suite d'un thread est extrêmement facile.	M	E	*	
62	The ability to follow threads may seem a small thing, but it's one of those small touches added in this version of Office that should pay big dividends in increased productivity.	La capacité de suivre des threads peut sembler une petite chose, mais c'est l'une de ces petites touches ajoutées dans cette version d'Office que devrait payer de gros dividendes dans l'accroissement de la productivité.	M	E	*	
63	There's a related feature that helps cut down on e-mail clutter -- the ability to "clean up" a conversation.	Il y a une fonctionnalité connexe qui aide à réduire le fouillis de courriel--la capacité à « nettoyer » une conversation.	M	E		
65	However, once you do that, all of the quoted and previous text and e-mails are	Cependant, une fois que vous faites cela, tout le texte entre guillemets et	M	E	*	
66	It would be more useful if you were given the option of hiding the text, not completely deleting it.	Il serait plus utile si vous ont été donnés l'option de cacher le texte, pas complètement la suppression.	M	E	*	
67	Also of note	Également de la note	M	E		
68	Microsoft also says that Outlook now includes Mail Tips that warn you against sending out e-mails that you perhaps shouldn't send.	Microsoft a également dit que Outlook inclut maintenant des conseils de messagerie que vous mettre en garde contre l'envoi des e-mails que vous ne	M	E	*	
69	So, for example, if you're sending a message to too large a group so that it seems as if its spam, or if you're sending mail to someone who is out of the office, or if you're sending an e-mail to external parties and doing so might compromise confidential information, you'll be warned.	Si, par exemple, si vous envoyez un message à un groupe de trop grand afin qu'il semble comme si son spam, ou si vous envoyez un courrier à une personne qui est hors du bureau, ou si vous désirez envoyer un e-mail à des tiers et cela risque de compromettre des informations confidentielles, vous serez averti.	M	E	1	?
70	It works only with Exchange, so I could not test this feature.	Il ne fonctionne qu'avec Exchange, donc je ne pourrais pas tester cette fonctionnalité.	M	E	*	
71	There's one thing that didn't change in this version of Outlook that many people wish had -- the exclusive use of Word as your e-mail text editor.	Il y a une chose qui n'a pas changé dans cette version d'Outlook que beaucoup de gens avaient--l'usage exclusif du mot comme éditeur de messagerie texte.	M	E	*	
76	There is, however, a free, third-party Outlook 2003/2007 add-in called Xobni	Il est, cependant, un libre, tiers Outlook 2003/2007 complément appelé Xobni	M	E		
77	A Xobni representative said the add-on will be updated to work with Outlook 2010 by the time it ships next year.	Un représentant de Xobni dit que l'Add-on sera mis à jour pour fonctionner avec Outlook 2010, au moment où qu'elle expédie l'an prochain.	M	E	*	
79	Minor, but useful, changes to Word	Des changements mineurs, mais utile, vers Word	M	E	*	
80	Word hasn't received nearly as significant changes as has Outlook.	Word n'a pas reçu presque comme des changements importants tout comme Outlook.	M	E	*	
81	But there are some tweaks.	Mais il y a quelques réglages.	M	E	*	
82	Search has gotten a nice boost with a set of features that let you search	Recherche a obtenu un coup de pouce sympa avec un ensemble de	M	E		

Figure 3: Bottom left-hand panel displays the Language Quality Issue data category values associated with the selected segment in the main editor window.

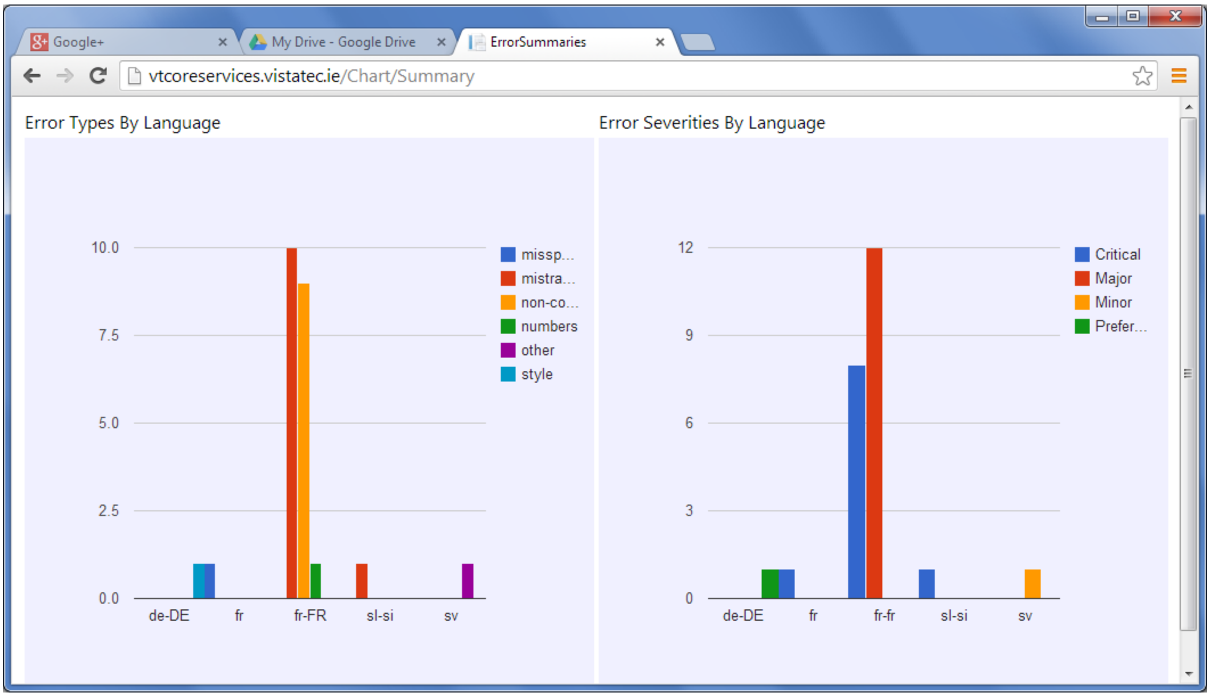


Figure 4: Quality Web Dashboard. LQI metadata can be sent via RESTFul API to VistaTEC endpoint and quality metrics immediately updated in portal dashboard.

5. Ocelot Open Source and Technical Details

VistaTEC's announced the availability of Ocelot as Open Source under the LGPL License at Localization World, Santa Clara in October 2013.

Source code is available at: https://github.com/ocelot

Documentation wiki is at: http://open.vistatec.com/ocelot

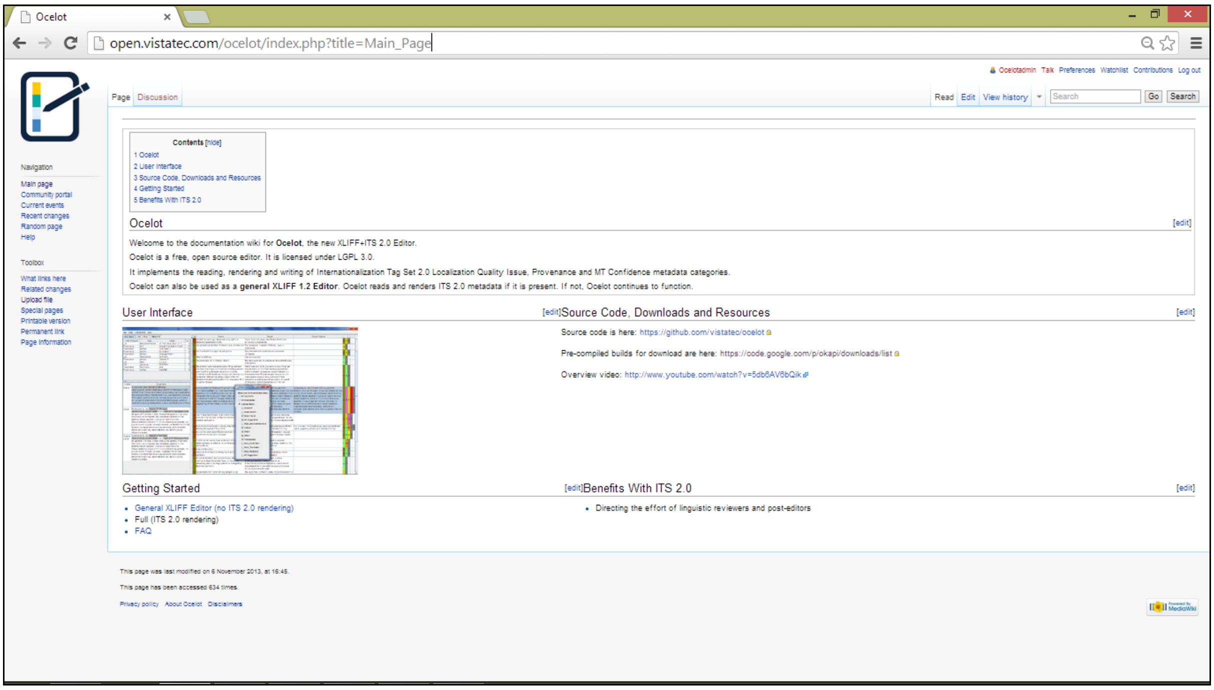
Binaries are available at:

Email list: okapitools@yahoogroups.com

Ocelot is written in Java.

It uses (and enhanced) several Okapi Framework classes, namely: `XLIFFFilter`, `XLIFFSkeletonWriter`, `ITSStandoffManager` and `ITSProvenance*`.

The easiest way of getting setup with the source code is to install the Oracle Netbeans IDE. Download the code from GitHub and open the Maven `pom.xml`. This will initiate Netbeans downloading all of the dependency libraries. Once this completes you can build and launch the application.



Contents [hide]
1 Ocelot
2 User Interface
3 Source Code, Downloads and Resources
4 Getting Started
5 Benefits With ITS 2.0

Ocelot [edit]

Welcome to the documentation wiki for **Ocelot**, the new XLIFF+ITS 2.0 Editor.

Ocelot is a free, open source editor. It is licensed under LGPL 3.0.

It implements the reading, rendering and writing of Internationalization Tag Set 2.0 Localization Quality Issue, Provenance and MT Confidence metadata categories.

Ocelot can also be used as a **general XLIFF 1.2 Editor**. Ocelot reads and renders ITS 2.0 metadata if it is present. If not, Ocelot continues to function.

User Interface [edit]

Source Code, Downloads and Resources [edit]

Source code is here: https://github.com/vistatec/ocelot

Pre-compiled builds for download are here: https://code.google.com/p/okapi/downloads/list

Overview video: http://www.youtube.com/watch?v=5db6AV6bQik

Getting Started [edit]

- General XLIFF Editor (no ITS 2.0 rendering)
- Full (ITS 2.0 rendering)
- FAQ

Benefits With ITS 2.0 [edit]

- Directing the effort of linguistic reviewers and post-editors

This page was last modified on 6 November 2013, at 16:45.

This page has been accessed 634 times.

Privacy policy About Ocelot Disclaimers

6. TANGIBLE OUTCOMES OF OCELOT

The technology startup Digital Linguistics (http://www.digitallinguistics.com) has adopted ITS 2.0 as its mechanism for writing its text analytics-generated Conformance Scores into XLIFF.

Ocelot and ITS 2.0 were presented at the TAUS Translation Quality Evaluation Summit held in Adobe on 9th October 2013. This event was attended by many of the world's top localizers.

Ocelot now forms the basis of a post-editing service within VistaTEC which is based fundamentally upon ITS 2.0 metadata.

VistaTEC's process now consists of the following steps:

1. Source files are converted to XLIFF 1.2. This can be done with Okapi Rainbow.

2. Files are sent to machine translation engine.

3. Files containing raw machine translation output are run through Okapi Checkmate. This adds ITS metadata for common errors.

4. Files are posted to Digital Linguistics' Review Sentinel service which adds "non-conformance" ITS markup.

5. Ocelot and a `rules.properties` file is distributed to post-editors. The `rules.properties` file contains rendering and segment filtering rules for 4x 10-point quality non-conformance bands. It also configures keyboard shortcut keys for adding Fluency/Adequacy scores.

6. Files are post-edited. Post-editors carry out edits using the 4 number 10-point non-conformance filters as their guide for segments to post-edit. Editors also add Fluency/Adequacy scores according to

proprietary guidelines.

7. Files are post-processed to harvest metadata added during post-editing.

8. Files are converted back to native format.