

# Ocelot XLIFF, metadata and enrichment

Phil Ritchie, Vistatec

Chase Tingley, Spartan Software Inc.



# The Motivation

- Innovative use of metadata
  - Low friction creation
  - Intuitive visualization
  - Interoperable serialization
  - High performant utilization
- To increase
  - Quality
  - Operational performance
  - Integration with other language technologies



# Introduction

Ocelot - AllFeatures.xlf

File View Filter Segment Extensions Help

Source Font Arial 12 Target Font Arial 12

Doc Stats LQI Prov Other ITS

Data Category	Type	Value	Count
Provenance	revPerson	phil.ritchie@vis...	2
Provenance	revOrg	Vistatec	2
LQI	style	1.0-2.0	2
LQI	misspelling	65.0-95.0	2
Provenance	person	olga.olaoghair...	6
Provenance	org	Vistatec	6
LQI	addition	75.0-75.0	1
LQI	non-conforma...	65.0-95.0	4
LQI	register	85.0-85.0	1
LQI	untranslated	95.0-95.0	1

Translations Concordance Search

Translation Results

#	Source	Target	Original Target
1	Missing from the Technical Preview is what will be a welcome change to Office in years -- a Web-based version for both enterprises and consumers.	Absent de la Technical Preview, c'est ce qui sera bienvenue changement au bureau au cours des années--une version basée sur le Web pour les entreprises et les consommateurs.	Project
2	Microsoft Office 2010, as revealed by the just-released Technical Preview, brings a set of important 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 et supplémentaires à la suite office de leader sur le marché.	
3	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--échelle du bureau, y compris des outils et une amélioration de la retouche photo coller opération.	
4	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.	Absent de la Technical Preview, c'est ce qui sera <b>bienvenue le plus important</b> changement au bureau au cours des années--une version basée sur le Web <b>pour</b> les entreprises et les consommateurs.
5	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.	
6	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.	
7	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.	
8	Global changes	Changements globaux	
9	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.	Office 2007 <b>a introduit le ruban, un changement majeur à une interface que de bureau remplacé le old les anciens menus and submenus with a graphical system that et sous-menus avec un système graphique que groups buttons for common tasks together in tabs de boutons pour les tâches courantes dans les onglets.</b>

Label Segments

Source: 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.

Target: 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.

Original Target: 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.

Edit Distance: 18

Filters

Show segments matching rules:

All Segments  All with metadata  Selected Rules:

- Critical non-conformance -
- Significant non-conformance \*
- Important non-conformance \*
- Possible non-conformance \*
- Fluency & Adequacy 4 4
- Fluency & Adequacy 3 3
- Fluency & Adequacy 2 2
- Fluency & Adequacy 1 1
- Quality Audit Critical Q
- Quality Audit Major Q
- Quality Audit Minor Q
- Quality Audit Preferential Q
- MT Segment M
- MTConfidence X

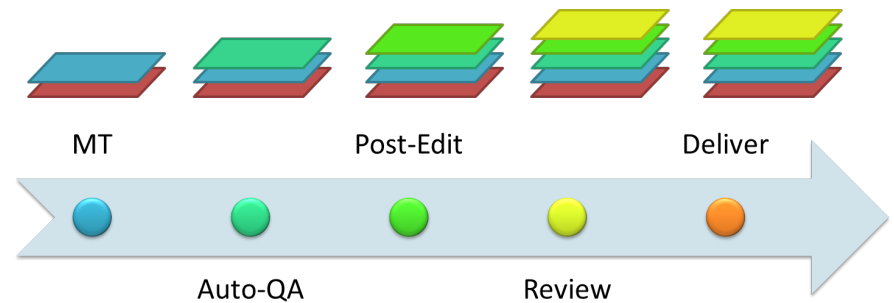
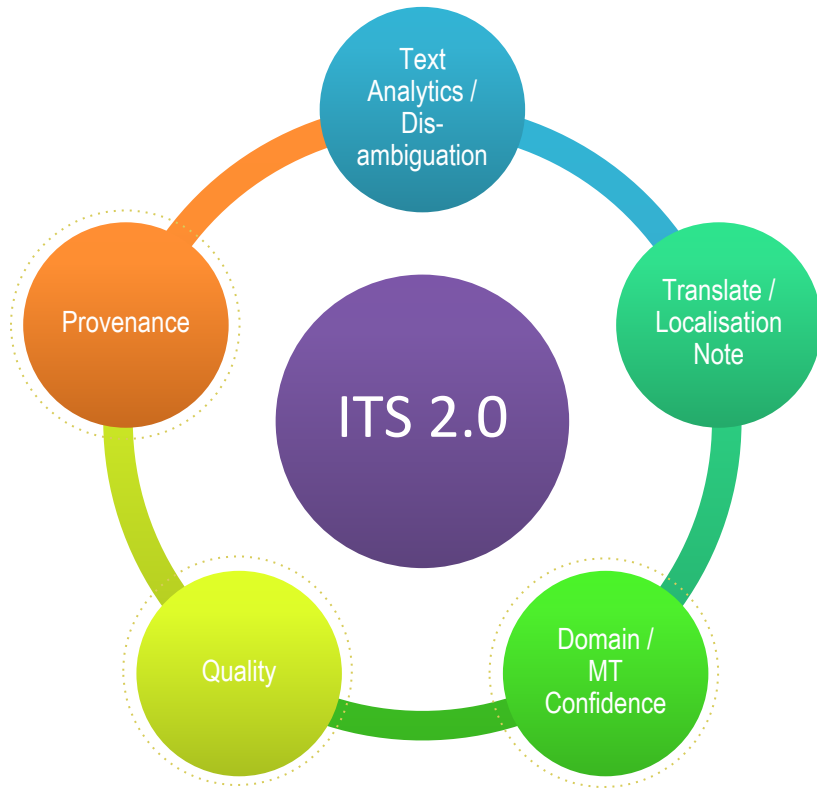
Filter segments by state-qualifier:

Show All  Show Only These:

- id-match █
- exact-match █
- fuzzy-match █
- mt-suggestion █



# Metadata the driving force



# The Rules File

- Allows users to describe rendering characteristics for:
  - ITS data category attributes and ranges of attribute values
  - XLIFF state-qualifier and phase-name

Filters

Show segments matching rules:

All Segments  All with metadata  Selected Rules:

<input type="checkbox"/> Critical non-conformance	*
<input type="checkbox"/> Significant non-conformance	*
<input type="checkbox"/> Important non-conformance	*
<input type="checkbox"/> Possible non-conformance	*
<input type="checkbox"/> Fluency & Adequacy 4	4
<input type="checkbox"/> Fluency & Adequacy 3	3
<input type="checkbox"/> Fluency & Adequacy 2	2
<input type="checkbox"/> Fluency & Adequacy 1	1
<input type="checkbox"/> Quality Audit Critical	Q
<input type="checkbox"/> Quality Audit Major	Q
<input type="checkbox"/> Quality Audit Minor	Q
<input type="checkbox"/> Quality Audit Preferential	Q
<input type="checkbox"/> MT Segment	M
<input type="checkbox"/> MT Confidence High	X
<input type="checkbox"/> MT Confidence Low	x

Filter segments by state-qualifier:

Show All  Show Only These:

<input type="checkbox"/> id-match	1
<input type="checkbox"/> exact-match	2
<input type="checkbox"/> fuzzy-match	3
<input type="checkbox"/> mt-suggestion	4



# Rules File Cameo



# Use Case Scenarios

- Configured to support characteristics of many Linguistic Quality and Machine Translation Assessment models
  - MQM, DQF, proprietary
  - Fluency and Adequacy
  - Using Quality Estimation
- Having designed flags metaphor, we found powerful techniques for using them...



# Efficient Metadata Capture Cameo





# ITS Data API

- Receives all LQI & Provenance data in a document as a series of callbacks
- Manually triggered (for now)
- Typical use case: external aggregation



# ITS Data API Cameo



# Segment Event API

- Callbacks for editing events
  - Session - (File Open, File Closed)
  - Segment - (Start Editing, Stop Editing)
- Initial goal: collection of timing information
- Additional scenarios
  - Validation/quality checks
  - Synchronization to an external application



# Segment Event API Cameo



# New Features (Ocelot 2.0)

- XLIFF 2.0 Support
  - Based on Okapi XLIFF 2.0 library
  - Caveat: storage of original targets
  - (Looking at Change Tracking module)
- TM Support
  - Leverage and concordance
  - Based on Okapi Pensieve backend
- UI Enhancements
  - Font selection
  - Customizable columns

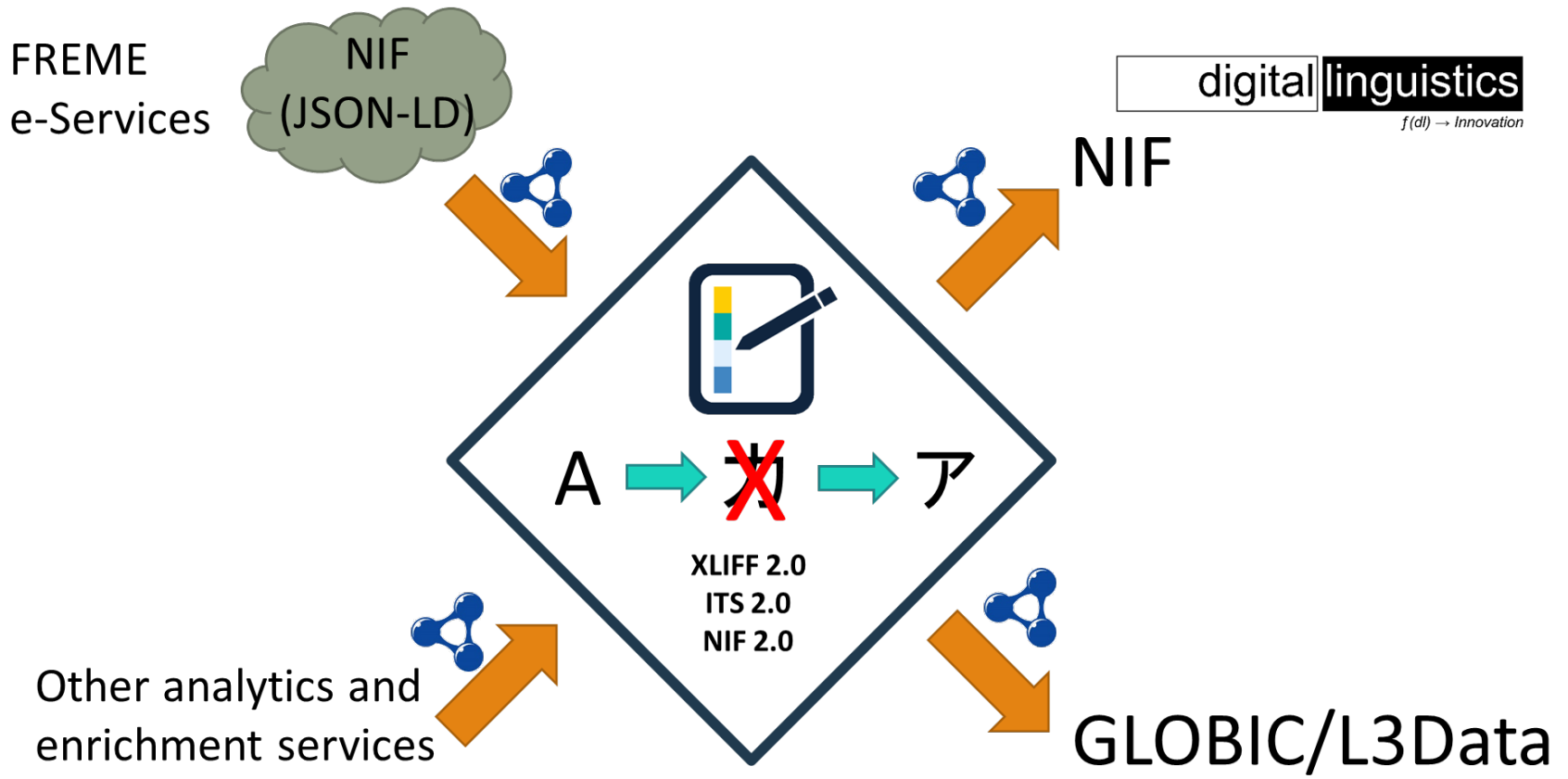


# Current Work & Future Plans

- Add features to support semantic enrichment and serialization
  - Working with Freme Consortium
  - <http://www.freme-project.eu>
- Provide interoperability with NLP Tools
  - NLP Interchange Format (NIF)
  - RDF and Linked Linguistic Data
- Integration with a wide range of services and repositories
  - Localization asset systems



# Vision



# Thank You

@philinthecloud, @ctatwork

