

L'extraction terminologique avec MemoQ 2015*

TriKonf 2015

*memoQ 7.8.54

Introduction

- Objectif

Compiler rapidement un glossaire bilingue avec l'outil d'extraction terminologique de memoQ.



Déroulement

- Principes de fonctionnement
- Avantages/Inconvénients
- Configuration
- Extraction des termes source
- Choix des termes pertinents
- Traduction des termes
- Exportation pour validation expert
- Importation dans Base Terminologique finale

Principes

- Fonctionnement statistique
 - Fréquence
 - Longueur
 - Mots vides
- Sources
 - Documents source
 - Mémoire de Traduction
 - Corpus (LiveDocs)
- Ressources et dictionnaires
 - Bases terminologiques
 - TM
 - LiveDocs
 - Dictionnaires...

Avantages/Inconvénients

Avantages

- Utilisation aisée
- Coût faible
- Réduction du temps de création

Inconvénients

- Bruit
- Silence

Pratique

- Prérequis :
 - Licence memoQ
 - Notions de bases de memoQ
 - Projet créé
 - Docs HTML (Wikipédia : TIC, réseaux...)
 - TM Microsoft (Windows XP)
 - TB Microsoft

Exécution

memoQ - Extraction terminologique TriKonf 2015

Prétraduire... Copier le texte source dans les segments cible vides... Remplir les segments ne contenant que des nombres Effacer la traduction... Rechercher... Remplacer... Verrouiller/déverrouiller les segments... Modifier l'état des segments... Contrôle qualité Modifications et conflits Module de balisage RegEx Extraire la terminologie... Confirmer et valider en MT Rechercher les abréviations Clear Comments

Traductions Chercher/remplacer État des segments

Accueil

Vue d'ensemble

Traductions

Ajoutez ou supprimez des documents, échangez des fichiers bilingues ou livrez votre travail

Bases LiveDocs

Mémoires de traduction

Bases terminologiques

Muses

Paramètres

Traductions

Documents Vues

rechercher

Structure Détails

Trier par nom

Nom	V	#	Progression
Domain Name System - Wikipedia, the free encyclopedia.html	1.0	8674	
Internet protocol suite - Wikipedia, the free encyclopedia.html	1.0	7362	
Network administrator - Wikipedia, the free encyclopedia.html	1.0	1768	
Network management - Wikipedia, the free encyclopedia.html	1.0	2118	

Extraire la terminologie...
Sessions d'extraction de terminologie...

Extraire la terminologie...
Garantir une parfaite cohérence terminologique en identifiant et en traduisant les expressions fréquentes avant le début du projet de traduction.

Projet (fre-FR) : 0% | Sélectionné : 0 documents | 0 mots | TR : 0 % | R1 : 0 % | R2 : 0 %

Paramètres

Extraire des candidats

Nom de session

Source

Documents Mémoires de traduction Bases LiveDocs

Tous les documents Tous les documents affichés

Documents sélectionnés Toutes les mémoires du projet Documents sélectionnés

MT principale MT sélectionnées

Options

Général

Longueur maximale (mots)

Nombre d'occurrences minimum

Séparateurs d'expressions
Rapport de longueur

Ignorer les mots contenant des nombres

Mots seuls

Longueur minimum (caractères)

Nombre d'occurrences minimum

Interrogation des bases terminologiques

Rechercher la traduction des candidats

Toutes les bases terminologiques du projet

Base terminologique de premier rang uniquement

Mots vides

Liste de mots vides

Mot	En début	À l'intérieur	À la fin

Mot

Mots vides

Extraire des candidats

Nom de session: Trikonf 2015 - Source documents

Source

Documents
 Mémoires de traduction
 Bases LiveDocs

Tous les documents
 Documents sélectionnés
 Toutes les mémoires du projet
 MT principale
 MT sélectionnées
 Tous les documents affichés
 Documents sélectionnés

Options

Général

Longueur maximale (mots): 4
Nombre d'occurrences minimum: 3
Séparateurs d'expressions: ;:!"'0[]?!\@>< > "" . .
Rapport de longueur: 1.50
 Ignorer les mots contenant des nombres

Mots seuls

Longueur minimum (caractères): 3
Nombre d'occurrences minimum: 3

Interrogation des bases terminologiques

Rechercher la traduction des candidats
 Toutes les bases terminologiques du projet
 Base terminologique de premier rang uniquement

Mots vides

Liste de mots vides: <nouvelle liste de mots vides> [registre] Enregistrer

Mot	À la fin	À la fin	À la fin
another	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
any	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
are	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
about	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
above	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Mot: Ajouter Supprimer sélection

OK Annuler Aide

Liste des candidats

memoQ - Extraction terminologique Trikonf 2015

memoQ | Projet | Documents | Préparation | Modifier | Affichage | Extraction de terminologie

Accepter comme terme | Ignorer terme | Sélectionner tous les segments | Activer/désactiver les résultats et la base terminologique | Fusionner les candidats | Annuler fusion des candidats | Fusionner préfixes & masquer | Masquer/afficher candidats plus courts | Ajouter aux mots vides | Rechercher termes maintenant | Exporter vers une base de terminologie | Exporter vers TaaS | Relancer la session

Terme | Candidats | Référence | Session

Accueil | Trikonf 2015 - Source documents x

Trier [Aucun raccourci] | Langue cible Français (France)

Filter [Ctrl+Maj+F] | Recherche uniquement dans les BT | Entrées acceptées fournissent des résultats de traduction

▼ \$	n°	M...	État	Source	Cible
58,00	69		Candidat	domain name System	
57,65	126		Candidat	domain name	nom de domaine
39,51	47		Candidat	Jump up ^	
39,15	242		Candidat	name	
37,21	230		Candidat	DNS	DNS
36,07	223		Candidat	Protocol	protocole
35,59	220		Candidat	management	
33,97	210		Candidat	domain	domaine
33,00	204		Candidat	network	réseau
32,49	71		Candidat	name System	
29,60	183		Candidat	Internet	internet; Internet

Occurrences Navigation : [Ctrl+Haut]/[Ctrl+Bas]

Domain Name System	DNS (Domain Name System)
Domain Name System	Domain Name System
Domain Name System resolver	résolution DNS
Domain Name System resolver	Domain Name System resolver

Source Microsoft Terminology (TB)

Résultats des bases terminologiques

system	Microsoft Terminology
system	Modifié le 31/08/2015 15:45 par gl12
System	
eng-US system	
fre-FR système	

Projet (fre-FR) : 0% | Ins | Éléments : 1824 | Acceptés : 0 | Candidats : 1824 | Masqués : 0 | Ignorés : 0

Traduire et valider

The screenshot displays the Trifoneo software interface for translating and validating terms. The main window shows a list of terms with columns for source and target. A context menu is open over the term "nom de domaine", with "Accepter comme terme" (Accept as term) highlighted. Below the list, the "Occurrences" section shows the translation of "domain namespace" to "espace de noms de domaine". The "Microsoft Terminology" window is also visible, showing the source "Microsoft Terminology" and the target "domain".

Source	Cible
domain names	nom de domaine
domain	
top-level domain	
domain name System	
Internationalized domain names	
Internationalized domain	
domain name System Security	
Load a random article	
domain names for applications	
fashion and adornment pictures	
cultural fashion and adornment	
top level domain	
domain names in applications	
fully qualified domain name	

Occurrences: Navigation: [Ctrl+Haut]/[Ctrl-Bas]

Source	Cible
domain namespace	domain namespace
domain namespace	espace de noms de domaine

Microsoft Terminology
Modifié le 31/08/2015 15:45 par gl12

eng-US domain
fre-FR domaine

Projet (fre-FR): 0% | Ins | Éléments: 1823 | Acceptés: 11 | Candidats: 1809 | Masqués: 0 | Ignorés: 3

Exportation

- Dans une BT
- D'une BT vers une feuille Excel pour validation

English_United_States	French_France
Internet Protocol	protocole Internet
Protocol	protocole
layer	couche
DNS	DNS
domain	domaine
domain names	nom de domaine
name server	serveurs de noms
System	système
TCP/IP	TCP/IP
Software	logiciel
top-level domain	domaine de premier niveau
network management	administration réseau

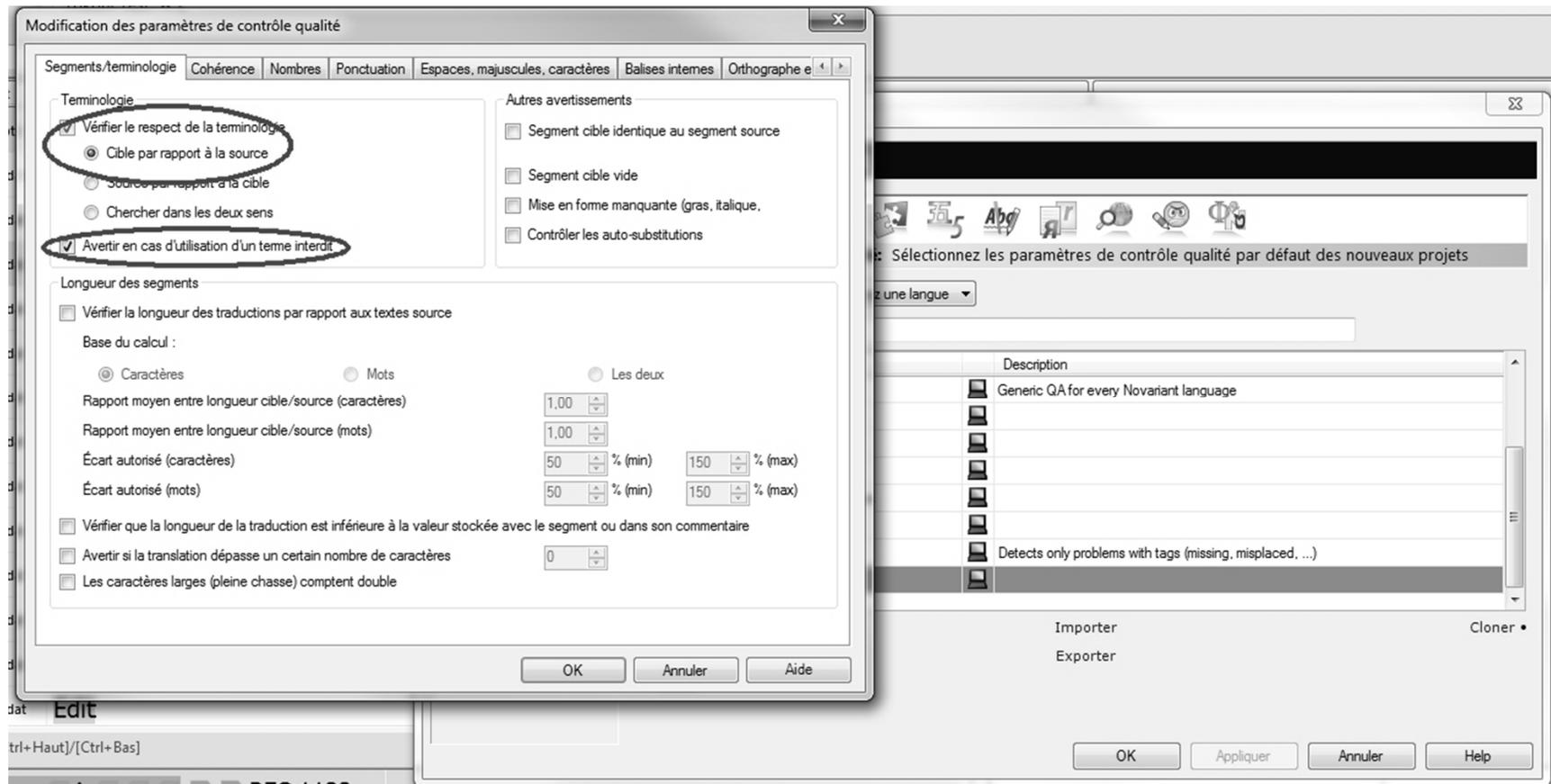
Importation

- Importation après relecture et confirmation

English_United_States	French_France
Internet Protocol	protocole Internet
Protocol	protocole
layer	couche
DNS	DNS
domain	domaine
domain names	nom de domaine
name server	serveurs de noms
System	système
TCP/IP	TCP/IP
Software	logiciel
top-level domain	domaine de premier niveau
network management	gestion de administration réseaux

- Création de la Base Terminologique définitive

Assurance qualité



Autres outils de gestion de la terminologie

- Extraction terminologique
 - Okapi Term Extractor – Gratuit, peu intuitif
 - Araya Bilingual Terminology Extraction – 800 €
 - SDL MultiTerm Extract – 400 €
- Gestion de la terminologie
 - SDL MultiTerm Desktop 2015 – 250 €
 - Xbench – Gratuit en évaluation

Questions ?

- Vos questions...



Merci pour votre attention.

Et encore une fois merci aux organisateurs
pour leur accueil !

Pour toute question : gabriel.lang@langnetfrance.fr

Le contenu de cette présentation sera en ligne sous forme de didacticiel :
<http://www.langnetfrance.com/blog/>