

Traducción y tecnologías: herramientas, procesos y recursos

Antoni Oliver

Introducción

Las tecnologías informáticas han introducido plenamente en el mundo de la traducción y los profesionales de la traducción contamos con un gran abanico de aplicaciones diseñadas para facilitar nuestro trabajo, mejorando el rendimiento y asegurando unos niveles de calidad óptimos. Estas tecnologías van desde la traducción asistida a la traducción automática, pasando por los sistemas de extracción y gestión de la terminología. El traductor, además, tiene a su disposición una gran cantidad de recursos lingüísticos: glosarios terminológicos, diccionarios y corpus paralelos que, empleados de manera adecuada pueden suponer un gran ahorro de tiempo.

Este libro presenta de una manera clara y con detalle todos los aspectos de las tecnologías aplicadas a la traducción. También enseña a encontrar, crear y gestionar recursos lingüísticos para la traducción, especialmente bases de datos terminológicas y memorias de traducción. El libro puede ser un manual de consulta para el traductor en activo y puede también servir como material de referencia principal en un curso universitario de grado, postgrado o máster sobre tecnologías aplicadas a la traducción.

El libro se complementa con una serie de enunciados de prácticas que se pueden descargar libremente de <http://lpg.uoc.edu/TIT>.

El autor del libro está abierto a recibir comentarios, críticas y sugerencias a través de su correo electrónico.

Antoni Oliver

aoliverg@uoc.edu

Noviembre de 2014

Índice

1. La traducción asistida por ordenador

- 1. Introducción
- 1.2. Componentes básicos de un sistema de traducción asistida
 - 1.2.a Entorno de trabajo
 - 1.2.b. Consulta a memorias de traducción
 - 1.2.c. Consulta a glosarios y diccionarios
 - 1.2.d. Combinación de traducción asistida y traducción automática
 - 1.2.e. Tratamiento de formatos
- 1.3. El proceso de traducción con un sistema de traducción asistida
 - 1.3.a. Tratamiento del formato
 - 1.3.b. Segmentación: el formato SRX
 - 1.3.c. Formatos de proyectos de traducción: el formato XLIFF
 - 1.3.d. Asignación de recursos a un proyecto de traducción
 - 1.3.e. Traducción
 - 1.3.f. Revisión dentro de la herramienta de traducción asistida
 - 1.3.g. Creación de los documentos traducidos
 - 1.3.h. Revisión en formato final
 - 1.3.i. Gestión de los recursos generados
- 1.4. Los sistemas de traducción asistida
- 1.5. Conclusiones
- 1.6. Para ampliar conocimientos
 - 1.6.1. Herramientas de traducción asistida on-line
 - 1.6.2. SRX y las expresiones regulares
 - 1.6.7. El corrector gramatical LanguageTool

2. Las memorias de traducción

- 2.1. Introducción
- 2.2. Indexación y recuperación de segmentos
 - 2.2.1. Indexación de memorias de traducción
 - 2.2.2. Cálculo de la similitud de segmentos
- 2.3. Coincidencia exacta y parcial
- 2.4. Combinación de unidades subsegmentales
- 2.5. Formato de intercambio de memorias de traducción: TMX
- 2.6. Creación de memorias de traducción
 - 2.6.a. Traducción con sistemas de traducción asistida
 - 2.6.b. Memorias de traducción y corpus paralelos
 - 2.6.c. Alineación manual de documentos
 - 2.6.d. Alineación automática de documentos
- 2.7. Memorias de traducción remotas compartidas y públicas
- 2.8. Trabajo con memorias de traducción
- 2.9. Análisis de proyectos y tarificación
- 2.8. Nuevas funcionalidades
 - 2.8.1. Sistema de traducción automática integrado
 - 2.8.2. Autocompletado inteligente
 - 2.8.3. Medidas de confianza
- 2.9. Conclusiones
- 2.10. Para ampliar conocimientos
 - 2.10.1. Corpus paralelos y memorias de traducción disponibles públicamente
 - 2.10.2. Etiquetadores morfosintácticos
 - 2.9.3. Propiedad de las memorias de traducción

3. Las bases de datos terminológicas

- 3.1. Introducción
- 3.2. Terminología y traducción
- 3.3. Bases de datos terminológicas, diccionarios generales y conocimiento enciclopédico
- 3.4. Búsqueda automática en bases de datos terminológicas
- 3.5. Formato de intercambio de bases de datos terminológicas: TBX
- 3.6. Creación de bases de datos terminológicas
 - 3.6.a. A medida que se traduce
 - 3.6.b. A partir de recursos terminológicos en Internet
 - 3.6.c. A partir de la Wikipedia
 - 3.6.d. Extracción automática de terminología
- 3.7. Extracción automática de terminología
 - 3.7.1. Definición
 - 3.7.2. Clasificación de métodos para la extracción automática de terminología
 - 3.7.3. Métodos estadísticos para la extracción automática de terminología
 - 3.7.4. Métodos lingüísticos para la extracción automática de terminología
 - 3.7.5. Medidas estadísticas
- 3.8. Conclusiones

4. La traducción automática

- 4.1. Introducción 1
- 4.2. Historia de la traducción automática 3
- 4.3. Tipos de sistemas de traducción automática 6
 - 4.3.1. Traducción directa 7
 - 4.3.2. Sistemas de transferencia 8
 - 4.3.3. Sistemas de interlingua 9
 - 4.3.4. Sistemas estadísticos 13
 - 4.3.5. Sistemas de traducción automática basada en ejemplos 19

4.3.6. Sistemas híbridos

4.4. Usos de la traducción automática

4.5. La traducción automática en Internet

4.6. Combinación de traducción automática y traducción asistida

4.7. Post-edición de traducción automática

4.8. Conclusiones

4.9. Para ampliar conocimientos

4.9.1. Historia de la traducción automática

4.9.2. Moses: un sistema de traducción automática estadística

4.9.3. Otros sistemas de traducción automática

5. Tratamiento de formatos

5.1. Introducción

5.2. Representación de la información textual: código de caracteres

5.2.a Sistemas de numeración

5.3. El lenguaje interno del ordenador y las unidades de medida en informática

5.3.a. Los múltiplos del byte

5.4. Representación de información no numérica

5.4.a. Representación de texto

5.5. Representación de la información textual: código de caracteres

5.5.a. Conceptos básicos

5.5.b. Algunas definiciones importantes

5.5.c. Los códigos de caracteres más habituales

5.5.d. Unicode

5.5.e. Detección de la codificación de caracteres

5.5.f. Cambio de la codificación de caracteres

5.6. La representación de la información no textual

5.6.a. Nombres de archivo y extensiones. Relación con el formato y la aplicación

5.6.b. El formato HTML

5.6.c. El XHTML

5.6.d. Open Document

5.6.e. Los formatos de documentos de Microsoft Word

Microsoft Word DOC (doc)

Microsoft Word 2003 XML (.xml)

Microsoft Word 2007/2010 XML (.docx)

5.6.f. El formato LaTeX

5.6.g. El formato DocBook

5.6.h. El formato PDF

5.7. XML

5.7.a. Introducción

5.7.b. Ejemplos sencillos de documentos XML

5.7.c. Estructura de los documentos XML

5.7.d. Los documentos XML bien formados

5.7.e. Tecnologías asociadas: XSLT y XPath

5.7.f. Editores de XML

5.7.g. Traducción de documentos XML

5.8. Los formatos XML utilizados en el mundo de la traducción

5.8.a. Intercambio de memorias de traducción: TMX

5.8.b. Intercambio de bases de datos terminológicas: TBX

5.8.c. Intercambio de proyectos de traducción: XLIFF

5.8.d. Intercambio de reglas de segmentación: SRX

5.8.e. Métricas GILT: GMX

6. Conclusiones

Para ampliar conocimientos

Tutoriales de W3Schools

Detección automática de lengua

Tablas de Unicode

Problemas de visualización de documentos relacionados con las fuentes

SC Unipad: Editor de Unicode