

XML-Bi: Procedimientos para la gestión de flujo documental multilingüe sobre XML/TEI

Joseba Abaitua, Arantza Domínguez, Carmen Isasi, José Luis Ramírez
Facultad de Filosofía y Letras, Universidad de Deusto
abaitua@fil.deusto.es

Inés Jacob, Idoia Madariaga
Facultad de Ingeniería, Universidad de Deusto

Arantza Casillas
Universidad del País Vasco – Euskal Herriko Unibertsitatea

Raquel Martínez
Universidad Rey Juan Carlos

Alberto Garay
Euskalgintza

Thomas Diedrich
Stella die KommunikationsFabrik

Organismo financiador: Departamento de Educación, Universidades e Investigación del Gobierno Vasco (PI1999-72)

Resumen: XML-Bi es un proyecto que tiene como principal objetivo el diseño de procedimientos de optimización del flujo documental en organismos de gran tamaño, poniendo especial énfasis en la subsanación de desfases en la traducción y publicación plurilingüe. Con este fin se está diseñando una metodología que permite la supervisión de las distintas fases del ciclo de vida de la documentación, desde el momento inicial de concepción hasta el final de distribución, pasando por todas las etapas intermedias de redacción, corrección, traducción, revisión y publicación. Este proceso incluye el reciclaje documental, es decir, la catalogación y recuperación para la reutilización de elementos aprovechables. El corpus documental se categoriza y almacena en formato TEI/XML, para facilitar la administración de los archivos, así como la consulta y recuperación por *web*.

1 Introducción

En la mayor parte de las empresas e instituciones no existe una definición precisa de los procedimientos que se deben aplicar a la

hora de elaborar y distribuir la documentación. La situación se agrava considerablemente cuando se requiere la publicación en varios idiomas, ya sea por motivos comerciales o legales. Todo ello conduce a que se produzcan frecuentes desajustes entre las versiones originales y traducidas, así como desfases y retrasos innecesarios entre la redacción original de los documentos y su distribución.

El proyecto XML-Bi contribuye a resolver este problema mediante la formulación de un conjunto de especificaciones que se traducen en procedimientos operativos de trabajo, así como en el desarrollo de recursos: corpus de referencia (LEGE-Bi), filtros para conversión de formatos (TEI/XML y TMX) y desarrollo de interfaces con la intranet, para la consulta y acceso al corpus (SARE-Bi).

2 Objetivos operativos

XML-Bi se desglosa de manera más específica en cuatro objetivos operativos:

1. Estudio de campo: Se han elegido cuatro departamentos de la Universidad de Deusto para realizar un estudio pormenorizado del flujo documental. Este estudio abarca cuestiones como la definición de los perfiles y tipología de los textos, el estudio de los

- requisitos de redactores y usuarios, los medios informáticos disponibles, la utilización de distintos soportes, el acceso y utilización de la intranet de la Universidad, cuantificación del déficit de traducciones, procesos de control de calidad, plazos y canales de distribución de la documentación.
2. Elaboración de especificaciones que describen de manera analítica los resultados del estudio de campo, de manera que permitan el diseño de procedimientos operativos para subsanar los problemas detectados.
 3. Desarrollo de procedimientos, recursos y herramientas según las especificaciones. Este objetivo se desglosa en cuatro módulos:
 - 3.1 Propuesta de procedimientos operativos de trabajo para el control del flujo documental multilingüe según normas de calidad EFQM.
 - 3.2 Compilación e instalación de corpus LEGE-Bi en intranet (SARE-Bi).
 - 3.3 Diseño y adaptación de segmentadores para corpus en formato TEI/XML y de filtros para conversión a TMX.
 - 3.4 Integración de recursos externos (lematizadores, glosarios terminológicos, etc.)
 4. Verificación y control de desarrollos.

3 Situación actual

El proyecto comenzó en enero de 2001 y tiene una duración de dos años. En la actualidad se está terminando la fase de estudio de campo y se ha comenzado la definición de las especificaciones. El corpus recogido abarca una tipología extensa de documentos, pero el número de muestras es reducido (entre 15 y 20 documentos por centro y tipo de documento), debido a la poca variedad que se da dentro de cada tipo.

Se ha elegido un modelo de servidor basado en arquitectura ZOPE. Con ello se pretende agilizar el acceso al corpus. Se han desarrollado filtros de TEI/XML y TMX para permitir la exportación a sistemas comerciales de memorias de traducción (DéjàVu, Transit o TradosTW).