

Xflow: Gestión de flujo de contenidos multilingües sobre XLIFF y TMX

Fernando Quintana Hernández

Índice general

1. Introducción	7
2. El proyecto	9
3. Workflow	11
3.1. El workflow en una organización	11
3.1.1. Las primeras soluciones	11
3.1.2. Los sistemas se refinan	12
3.1.3. Tipos de workflow	13
3.2. Un aplicación de workflow	13
3.2.1. Algunas definiciones	14
3.2.2. Características de una aplicación de workflow	14
3.2.3. Tipos de herramientas	15
3.2.4. Tipos de aplicaciones de workflow	16
4. Workflow de localización	17
4.1. Internacionalización y localización	17
4.2. La solución de localización adoptada por CodeSyntax	17
4.2.1. Zope y el producto Localizer	17
4.2.2. ¿En qué consiste Localizer?	18
4.3. El workflow de localización en CodeSyntax	19
4.3.1. Los mensajes del interfaz de usuario	19
4.3.2. Los contenidos	20
4.3.3. Procedimientos susceptibles de mejoras	20
4.4. Las personas y los roles	21
4.4.1. Los programadores	21
4.4.2. Los editores contenidos	21
4.4.3. Los traductores	21
4.4.4. El revisor	24
4.4.5. Otros	26
5. Fases de un documento multilingüe	29
5.1. Los estados de un documento	30

6. El producto X-Flow	33
7. Aportación de X-Flow	37

Índice de figuras

4.1. Pantalla: Tarea de edición de contenidos	22
4.2. Pantalla: Petición de traducción de contenidos (1)	23
4.3. Pantalla: Petición de traducción de contenidos (2)	24
4.4. Pantalla: Tarea de traducción	25
4.5. Pantalla: Tarea de revisión	26
5.1. Estados de un documento.	30
5.2. Estructura de un documento.	31
6.1. Diagrama de clases	35

Capítulo 1

Introducción

El objetivo general del proyecto X-Flow es facilitar la gestión del flujo (Fisher, 2003) de contenidos multilingües en las distintas fases de desarrollo de los proyectos de traducción y localización (Savourel, 2001), con el fin de optimizar los recursos destinados a tales proyectos, especialmente en los sistemas de información y publicación en Internet.

La localización es el proceso por el cual se adapta una aplicación a los requerimientos locales de una determinada región (Palomar, 2002). En este proceso intervienen varios agentes en el papel de programadores, administradores de contenidos y traductores. En un entorno de publicación de contenidos multilingües en Internet, se genera un nutrido flujo de documentos multilingües que van pasando de un agente a otro, atravesando diferentes fases hasta completar su ciclo de vida.

En el mundo de la gestión documental multilingüe y su publicación en Internet, la comunidad Zope (Zope Community, 2003) ha alcanzado un gran protagonismo. Dentro de esta comunidad, el módulo Localizer (Palomar, 2003) es una solución adecuada al problema de la localización de los contenidos web.

Por otro lado en los últimos años han ido apareciendo una serie de estándares en el horizonte de la localización. XLIFF¹ (Karasek, 2001) nace para recoger el testigo de TMX², incorporando facilidades para la gestión de flujo o el control de versiones.

El proyecto X-Flow pretende aportar una solución adecuada que complementa a los productos como Localizer en la gestión del flujo de documentos multilingües, basándose en los estándares de intercambio TMX y XLIFF.

¹XML-based Localisation Interchange File (OASIS XLIFF TC, 2003)

²Translation Memory eXchange (Localization Industry Standards Association (LISA), 2002; Josu Gómez, 2001)

Capítulo 2

El proyecto

Las tareas que se llevan a cabo durante el desarrollo del proyecto se reparten en varias fases:

1. Estudio pormenorizado de los mecanismos de desarrollo de aplicaciones web multilingües que se utilizan en la empresa CodeSyntax, analizando las herramientas software, los métodos de traducción de textos y el flujo de información que se produce.
2. Estudio analítico de las herramientas desarrolladas con anterioridad por el equipo investigador y las utilizadas actualmente por la empresa CodeSyntax; definición de las condiciones de implantación de los estándares TMX y XLIFF y selección del entorno y herramientas de programación.
3. Elaboración de las especificaciones del desarrollo del proyecto, describiendo los objetivos específicos, requisitos del sistema, características de la solución y planificación de la integración, mecanismo de prueba y validación.
4. Desarrollo de la solución. Adaptación de las herramientas que vayan a reutilizarse, implementación de programas específicos, distribución de las herramientas desarrolladas y elaboración de manuales de referencia.
5. Implantación de la solución en la empresa CodeSyntax y posterior evaluación.

Capítulo 3

Workflow

3.1. El workflow en una organización

Es común en las organizaciones que un sinnúmero de documentos circulen de un sitio para otro dentro de un grupo de trabajo (Elmagarmid y DU, 1998; David Hollingsworth, 1995; Simon y Marion, 1996). Estos documentos tienen una serie de características en común :

- a. Son creados por un usuario dado.
- b. Se envían al siguiente usuario que es la autoridad revisora¹.
- c. Se envían a distintas entidades con autoridad para aprobarlos², quienes pueden aprobarlo en paralelo.
- d. Se envían a una autoridad de aprobación final.
- e. Se envían al departamento donde finalmente será utilizado para lo que se diseñó.

3.1.1. Las primeras soluciones

Inicialmente, las organizaciones que trataban de controlar su workflow utilizaban elementos como una base de datos compartida, unos flags o semáforos, unas variables de estado y unos procesos que chequean los flags y datos asociados.

La secuencia básica que se daba era la siguiente:

1. Los usuarios de una aplicación introducen unos datos.
2. Estos datos se almacenan en una base de datos compartida.

¹Reviewing authority

²approving authorities

3. Un proceso evalúa los flags y hace que cuando la autoridad de aprobación correspondiente se conecte al sistema sea avisado de que tiene que estudiar esos datos.
4. La secuencia continúa a través de todas las autoridades de autorización y destinatarios de los documentos o fragmentos de documentos.

3.1.2. Los sistemas se refinan

A principios de los 90, animados por el crecimiento de la computación distribuida, comenzó una gran discusión sobre las características que un sistema de workflow debía tener. Para 1993 ya existía un conjunto de principios que deberían poderse hacer en un entorno de workflow. Estas características podían ser las siguientes:

1. Soporte para computación distribuida. La tecnología existente en este campo debe reutilizarse para los entornos de workflow. La comunicación entre redes, el paradigma cliente-servidor, el middleware, el proceso de transacciones complejas, etc, son servicios básicos que son validos para un entorno de workflow.
2. Las reglas de negocio deben estar soportadas por el sistema de workflow. Hay varios tipos de reglas que deberían estar representadas por el motor de workflow:
 - a. Quién está involucrado. En todos los procesos suele haber personas involucradas. (creadores de información, revisores, autoridades de aprobación, etc...) El sujeto puede ser una persona física (fulanito pi) o un cargo dentro de una organización (el asesor de contenidos). También puede que haya una segunda persona (física o no) que haga el papel de la primera en caso excepcional.
 - b. El elemento tiempo. Todos los procesos de una organización suelen tener la dimensión del tiempo incluso cuando este vale "no importa".
 - c. Las dependencias. Unos procesos dependen de otros o requieren información de otros procesos para continuar. Estas dependencias deben ser controladas por el sistema de workflow con la apropiada gestión de los elementos quién y cuando.
 - d. Sustituciones. No solo en el sentido de una persona que hace el trabajo de otra, sino la aparición de caminos alternativos o condicionados por alguna circunstancia.
 - e. Evolución. Las reglas de negocio de una organización pueden cambiar a lo largo del tiempo. El sistema de workflow debe ser modificable para adaptarse a las características de la organización.

- f. Privilegios de información. El motor de workflow debe ser capaz de soportar conceptos como la privacidad y la seguridad de los datos

3.1.3. Tipos de workflow

La naturaleza del workflow de una organización puede ser de varios tipos en función de los trabajos que se realicen en ella:

1. Administrativo. Hay un flujo de documentos no demasiado voluminoso. Los procesos suelen estar basados en la cumplimentación de formularios. Los procesos que se automatizan no son los críticos de la organización y suelen representar un bajo porcentaje del trabajo total de la organización.
2. Producción. Los procesos automatizados son los principales de la organización. El flujo que se maneja suele ser muy grande y sujeto a estrictas reglas de procesamiento y condiciones temporales.
3. Ad-hoc. Los workflow ad-hoc resolverán problemas desconocidos a priori y suelen aplicarse en forma de proyectos. Las reglas que rigen los procesos suelen ser creadas dinámicamente. Un workflow ad-hoc pone énfasis en que las tareas no se queden sin hacer o procurar la solución óptima en lugar de buscar una mejora considerable de los procesos críticos de la organización.

3.2. Un aplicación de workflow

Un sistema de workflow no consiste en el mero envío de emails a distintas personas recordándoles sus tareas. Sin embargo, el email puede ser un importante elemento en un sistema de workflow como principal mecanismo de comunicación de alertas a distintos usuarios.

El sistema intenta asegurar que las tareas se realizan bien y se realizan en el orden correcto. No se deben negar ni ocultar los problemas.

Los distintos sistemas de workflow pueden ser muy específicos para adaptarse a un dominio particular, sin embargo, el motor de workflow en sí debe permitir la presentación y administración de las características descritas anteriormente (usuarios, procesos, tiempos, alternativas, etc).

En un sistema de workflow es importante la parte que gestiona las alarmas y la resolución de contratiempos por las personas correspondientes y en los tiempos indicados.

3.2.1. Algunas definiciones

Un proceso de workflow³ es un conjunto, paralelo o en serie, de actividades⁴ coordinadas, conectadas para conseguir un objetivo común. Una actividad de proceso se define como un paso lógico o descripción de una pieza de trabajo que contribuye a completar el objetivo propuesto. Esta actividad puede ser manual o automatizada.

En primer lugar es necesario especificar un proceso de workflow a través de un lenguaje de definición de proceso⁵. Posteriormente el proceso es ejecutado por un sistema de administración de workflow⁶.

Un sistema de administración de workflow es un sistema que define completamente, administra y ejecuta procesos de workflow mediante la ejecución de software en un orden determinado por una representación electrónica del proceso de workflow.

3.2.2. Características de una aplicación de workflow

Existen varios rasgos que pueden definir una aplicación. Para definir concretamente una solución se deben estudiar las siguientes características:

- a. Representación gráfica. Es posible representar los procesos de workflow como mapas gráficos.

Pueden aparecer distintos elementos del workflow mediante iconos y líneas uniendo iconos.

Se pueden representar todos los elementos en un solo nivel o en varios niveles de detalle.

Se puede añadir una descripción de lo que se hace en cada paso.

- b. La arquitectura. La mayoría de sistemas de workflow utilizan el paradigma cliente/servidor. Suele haber un motor de workflow que actúa como coordinador y almacena meta-información en una base de datos subyacente. Otros componentes actúan de clientes del motor de workflow como el monitor de procesos, el iniciador de procesos o el controlador de procesos.

- c. El modelo de datos. Un sistema de workflow manipula varios tipos de workflow:

Los datos de control de procesos solamente los utiliza el motor de workflow.

³Process workflow

⁴Process activities

⁵Process definition language

⁶Workflow management system (WFMS)

Los datos relevantes de procesos son utilizados por el motor de workflow para determinar las transiciones de estado de los procesos de workflow.

- d. El modelo de usuario. Es importante separar el concepto de rol lógico y persona física.

Un rol lógico es la especificación de un conjunto de capacidades necesarias para realizar alguna tarea. Esto se suele definir en tiempo de diseño.

Las personas físicas que poseen un rol, cuentan con el correspondiente conjunto de capacidades.

- e. El grado de complejidad de las reglas. El sistema debe permitir la ejecución de procesos de workflow más complicados que una simple secuencia.
- f. Herramientas. Un conjunto de herramientas adicionales son de gran utilidad. Estas pueden cumplir diferentes propósitos:

Monitorizar procesos.

Tracear procesos.

Controlar procesos.

Desarrollo de procesos.

Animación y simulación.

- g. El sistema de workflow debe automatizar todos los procesos posibles y debe hacerlos transparentes al usuario.

3.2.3. Tipos de herramientas

En distintos escenarios de aplicación de un sistema de workflow, se utilizarán distintas herramientas. Por ejemplo, en un entorno orientado a la elaboración, aprobación y almacenaje de documentos se utilizarían las siguientes aplicaciones:

- a. Una herramienta de edición de textos.
- b. Herramientas de marcado (Markup tools) para el proceso de marcado.
- c. Una base de datos orientada a objetos para la fase de almacenamiento y acceso.
- d. Un servicio de directorio donde publicar los documentos y hacerlos accesibles por toda la organización.

- e. El correo electrónico, para las notificaciones y las comunicaciones privadas.
- f. Herramientas de Trabajo en grupo (Groupware) para mantener discusiones entre varias personas.
- g. Otras herramientas apropiadas.

3.2.4. Tipos de aplicaciones de workflow

A la hora de abordar un problema de workflow debe decidirse el tipo de enfoque que se le va a dar a la solución. Existen distintos tipos de aplicaciones de workflow:

- a. Aplicación para un dominio particular. Las aplicaciones de este tipo dan solución a un problema concreto y no son reutilizables para abordar otros tipos de workflow.
- b. Componentes workflow. Implementan una funcionalidad parcial de un sistema de administración de workflow. Necesita otros componentes que cubran el resto de las funcionalidades.
- c. Motores workflow⁷ abiertos. La especificación de los procesos de workflow se introducen mediante una determinada interfaz. Esta especificación es interpretada por el motor de workflow.
- d. Middleware con servicios de workflow.

⁷Workflow engine

Capítulo 4

Workflow de localización

4.1. Internacionalización y localización

Cuando se desarrollan sitios web con contenidos dinámicos multilingües es necesario tener en cuenta varios aspectos. Uno de ellos es la internacionalización de las aplicaciones web desarrolladas, es decir, es necesario que cada aplicación sea capaz de mostrarse en diferentes idiomas.

También es necesario adaptar cada sitio web a los requerimientos locales de determinadas regiones; por ejemplo, hay que proporcionar la traducción de los mensajes en el interfaz, hay que especificar en qué formato se debe mostrar la fecha o las cantidades monetarias, etc. Este último proceso es el de localización de la aplicación web. La internacionalización de una aplicación se realiza una sola vez; en el mejor de los casos durante el desarrollo de la misma. La localización se realiza una vez por cada idioma que se desee soportar.

4.2. La solución de localización adoptada por CodeSyntax

4.2.1. Zope y el producto Localizer

CodeSyntax se ha encontrado con esta problemática desde el principio, puesto que la gran mayoría de las aplicaciones web que desarrolla necesitan estar internacionalizadas y es preciso localizarlas para que la gestión de los contenidos multilingües sea mucho más sencilla. La solución a esta problemática aparece dentro de la comunidad Zope. Existen varios productos para llevar a cabo la internacionalización y localización de las aplicaciones elaboradas sobre Zope. Localizer es uno de los productos más utilizados para este propósito.

Localizer tiene una filosofía similar a la de la herramienta GNU Gettext (Free Software Foundation, 2002). Con esta herramienta, la internacionalización de una aplicación se lleva a cabo introduciendo determinadas marcas

especiales en el código fuente, indicando los mensajes que deben ser traducidos. En el proceso de localización es el siguiente:

A través de un programa de línea de comandos, se analiza el código fuente y se genera un fichero PO por cada idioma involucrado. Cada fichero contiene los mensajes a traducir.

Estos ficheros PO son editados por traductores humanos con un editor de texto cualquiera que introducen la traducción de cada mensaje.

En último lugar, los ficheros PO se compilan para generar ficheros binarios llamados MO, que se utilizan en tiempo de ejecución.

Las marcas especiales introducidas en el código fuente son en realidad llamadas a función que reciben como parámetro el mensaje a traducir y devuelven su traducción, extraída de los ficheros MO. Localizer aprovecha de este estándar de facto en la comunidad del software libre, tanto la experiencia como las herramientas desarrolladas en torno a Gettext y a los ficheros PO, (editores, etc...).

4.2.2. ¿En qué consiste Localizer?

Zope proporciona un interfaz web a través del cual pueden construirse aplicaciones. A través de estas pantallas de gestión es posible manipular los diferentes tipos de objetos que ofrece Zope (carpetas, imágenes, ficheros, etc...). La mayoría de los productos disponibles para Zope, proporcionan nuevos tipos de objetos con sus respectivas pantallas de gestión.

Localizer añade cuatro tipos de objetos para realizar la internacionalización de un sitio web, gestionar contenidos multilingües o para definir la política de negociación de idioma:

El primer objeto es el catálogo de mensajes (MessageCatalog). Es una base de datos donde se almacenan mensajes y sus traducciones a distintos idiomas. Estos mensajes suelen ser textos cortos del interfaz de usuario, por ejemplo: la etiqueta de un botón, el texto de un enlace, etc... Este objeto tiene un interfaz web a través del cual se gestionan y mantienen estas traducciones. También se puede acceder a esta base de datos a través de un interfaz similar al Gettext de GNU, es decir, llamando a una función con el mensaje como parámetro, esperando recibir la traducción como retorno. Los contenidos de esta base de datos también se pueden exportar e importar a y desde un fichero PO al estilo de la herramienta GNU.

Otro tipo de objetos es el contenido local (LocalContent). En este objeto se recogen contenidos multilingües, es decir, el texto original y sus traducciones a otros idiomas. Con este objeto también es posible

indexar los contenidos de cada idioma por separado. A través de un solo interfaz web, es posible editar cualquier versión de los contenidos, es decir, en una misma pantalla se muestra simultáneamente el texto en todos los idiomas. Cuando se visita el sitio web, Localizer se encarga de mostrar el texto en el idioma adecuado.

El objeto carpeta local (Local Folder) es una solución genérica para cualquier otro caso de localización. En este objeto se pueden almacenar diferentes versiones de cualquier objeto (una por cada idioma). La carpeta local sirve para localizar cualquier tipo de objeto binario como imágenes, documentos sonoros, etc... .

Por último, Localizer lleva a cabo la negociación del idioma. Esto significa que cuando el usuario accede a una página web, Localizer decide en qué idioma servirla. Por defecto esta negociación del idioma se basa en un solo criterio que es la cabecera `Accept-Language` del protocolo HTTP. Sin embargo existe otro tipo de objeto llamado Localizer (del mismo nombre que el producto), que permite personalizar la negociación del idioma. Esta clase incluye varios métodos para realizar tareas tales como escribir o en una cookie el idioma preferido del usuario, especificar el idioma en la url, modificar el idioma por defecto, etc... .

4.3. El workflow de localización en CodeSyntax

La empresa CodeSyntax encuentra una ventaja competitiva en la utilización de las herramientas anteriormente comentadas. La combinación de Zope y Localizer agiliza el desarrollo y mantenimiento de sitios web multilingües.

CodeSyntax pone un especial interés en que las aplicaciones web que desarrolla estén internacionalizadas. En la creación de sitios web se hace uso de las facilidades `MessageCatalog`, `LocalContent`, `LocalFolder` y `Localizer` que proporciona el producto Localizer de forma que el esfuerzo de localización es pequeño.

4.3.1. Los mensajes del interfaz de usuario

Los textos que aparecen en el interfaz de usuario son introducidos en un idioma inicial y la traducción a otros idiomas es llevada a cabo por el propio personal desarrollador. En ocasiones, cuando la dificultad de la traducción es alta, estos textos son remitidos a un traductor profesional a través del correo electrónico, como si se tratase de texto perteneciente a contenidos en general.

4.3.2. Los contenidos

Los contenidos que se muestran en un sitio web se introducen en un determinado idioma. La persona que se encarga de administrar y mantener el sitio web, es quien ha de remitir estos contenidos a un traductor profesional para ser traducidos al idioma correspondiente. El administrador accede al objeto Zope que almacena el texto en un determinado idioma y copia el texto a un mensaje de correo electrónico. El correo electrónico es enviado al traductor adecuado, es decir, que conozca el idioma original y el final.

El traductor realiza la traducción del texto sin conocer el contexto más allá de lo indicado en el mensaje electrónico y responde al administrador adjuntando la traducción en el mensaje. Una vez recibida la respuesta, el administrador copia el texto traducido y lo introduce en el objeto Zope correspondiente.

Es posible también que el traductor acceda directamente al objeto Zope y realice la traducción a través del interfaz web que Localizer proporciona.

4.3.3. Procedimientos susceptibles de mejoras

En este flujo de documentos, están involucradas varias personas. El desarrollador, la persona que administra y mantiene los contenidos del sitio web y los traductores de los distintos idiomas soportados.

No es posible mantener sincronizados todos los contenidos del sitio web, es decir, cuando la persona que mantiene los contenidos en la web modifica una parte de un texto, los traductores a otros idiomas no conocen ese cambio hasta que el administrador no les remite un nuevo texto a traducir. Por lo tanto no se hace un control de versiones que indique que documentos necesitan ser revisados o cuales están preparados para ser publicados.

Los traductores no cuentan con un interfaz personalizado, de acuerdo a un perfil de "traductor que conoce unos idiomas determinados", es decir, un traductor de euskera a francés solo puede acceder a un interfaz web donde se le muestran los contenidos en castellano, euskera, inglés, alemán, francés, etc. El interfaz web ofrece más información de la que necesitan.

A la hora de realizar traducciones de forma colaboradora, no existe una gestión de la concurrencia para permitir la edición simultanea de documentos.

Las traducciones realizadas a través de este interfaz web no están asistidas; el sistema no utiliza información extraída de traducciones ya realizadas para sugerir al traductor posibles soluciones, es decir, no se hace uso de las memorias de traducción ni del estándar TMX. De la misma forma no es posible intercambiar con otros sistemas la información extraída de las traducciones realizadas como podría hacerse utilizando el estándar XLIFF.

4.4. Las personas y los roles

Las personas que van a utilizar el sistema podrían agruparse en varios perfiles. De esta forma podríamos hablar de programadores, editores de contenidos, traductores, revisores de contenidos y usuarios en general.

Cada persona que se conecta al sistema debe hacerlo a través de un interfaz de usuario adecuado a sus necesidades. Por ejemplo, un traductor que se conecta al sistema debe encontrarse con las tareas de traducción relativas a sus capacidades (los idiomas que conozca) o su competencia (grado de confidencialidad, publicabilidad, etc...)

4.4.1. Los programadores

Las personas que forman el grupo de programadores son profesionales que tiene conocimientos en lenguajes de programación como Python, Dtml, Tal, JavaScript, xhtml, etc... Estos serán los encargados de configurar y mantener el sistema de gestión de flujo.

Un programador debe contar con una API mediante la cual consulte si un documento está en un estado “publicable” o no. Un programador que utilice las propiedades locales de un objeto multilingüe o haga una llamada al catálogo de mensajes deberá contar con un mecanismo que le devuelva el estado del objeto. También debe poder acceder a las distintas versiones del documento.

4.4.2. Los editores contenidos

Los editores de contenidos son personas sin profundos conocimientos técnicos. Tal vez tengan algunas nociones de xhtml y JavaScript. La labor de estos personajes es introducir y mantener los contenidos de un sitio web. También se encargan de escribir los mensajes del interfaz de usuario.

Cuando el editor de contenidos accede al sistema puede encontrarse en distintas pantallas en función de la naturaleza de los contenidos que vaya a editar. Cuando el texto que se va a introducir es un mensaje corto para el interfaz de usuario, la pantalla que se encontrará será la de administración de mensaje, en el catalogo de mensajes (MessageCatalog). Si el texto es más extenso, la pantalla correspondiente será la de administración de las propiedades locales (ver figuras 4.1, 4.2 y 4.3) de un objeto multilingüe (BilEduki).

4.4.3. Los traductores

Los traductores son los encargados de traducir los contenidos y los mensajes de interfaz de un idioma a otro. Estas personas dominarán dos o más idiomas.

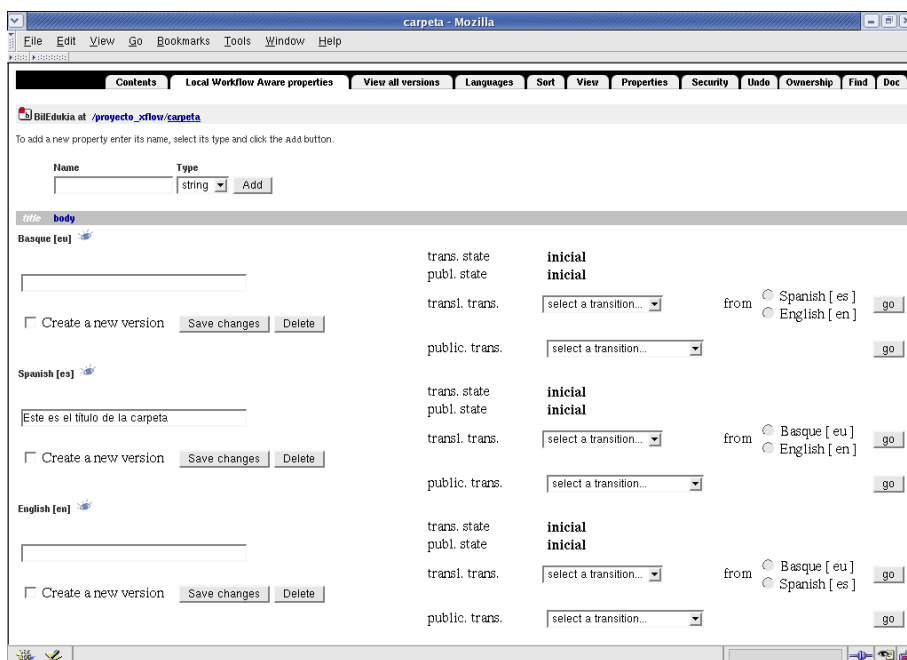


Figura 4.1: Pantalla: Tarea de edición de contenidos

El traductor no utilizará el sistema como una solución CAT¹ completa. Una gran parte de la interacción del traductor con el sistema se basa en la técnica de copiar/pegar. El traductor deberá contar con varias pantallas para realizar diferentes tareas:

La pantalla que se va a encontrar un traductor al entrar en el sistema mostrará información sobre las traducciones pendientes y las realizadas por el traductor anteriormente.

- a. Una lista de tareas de traducción pendientes con la siguiente información:
 - I. Quién ha creado la tarea de traducción,
 - II. Fecha de creación,
 - III. Prioridad,
 - IV. Número de palabras,
 - v. Título de la traducción (primera línea, etc.),
 - VI. Tipo de traducción (Hacer la traducción / Revisar la traducción)
 - VII. Categoría del documento (o si se trata de un mensaje),
 - VIII. Idioma origen,

¹Computer Aided Translation

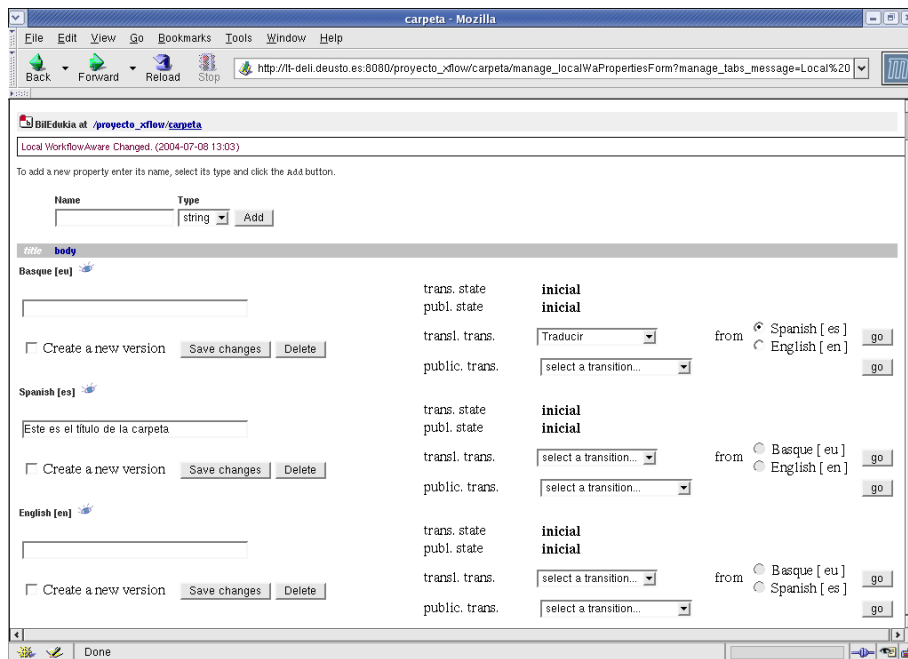


Figura 4.2: Pantalla: Petición de traducción de contenidos (1)

- IX. Estado del idioma origen
 - X. Idioma destino,
 - XI. Estado del idioma destino
 - XII. Estado en que se encuentra este documento en otros idiomas.
- b. Una serie de controles permitirán filtrar el listado de las tareas u ordenarlas en función de algún criterio.

Cuando el traductor selecciona una tarea de traducción, se encontrará con una pantalla (ver figura 4.4) en la que podrá realizar la traducción. Es importante mantener el contexto del texto a traducir por lo que se mostrará el documento completo en el que se encuentre el texto a traducir. En esta pantalla aparecerán varios elementos:

- a. Mediante pestañas (tabs) u otros controles se podrán acceder a las distintas versiones de este documento.
- b. El texto del documento en los idiomas origen y destino alineados.
- c. El texto del documento en los idiomas origen y destino en forma de texto corrido.

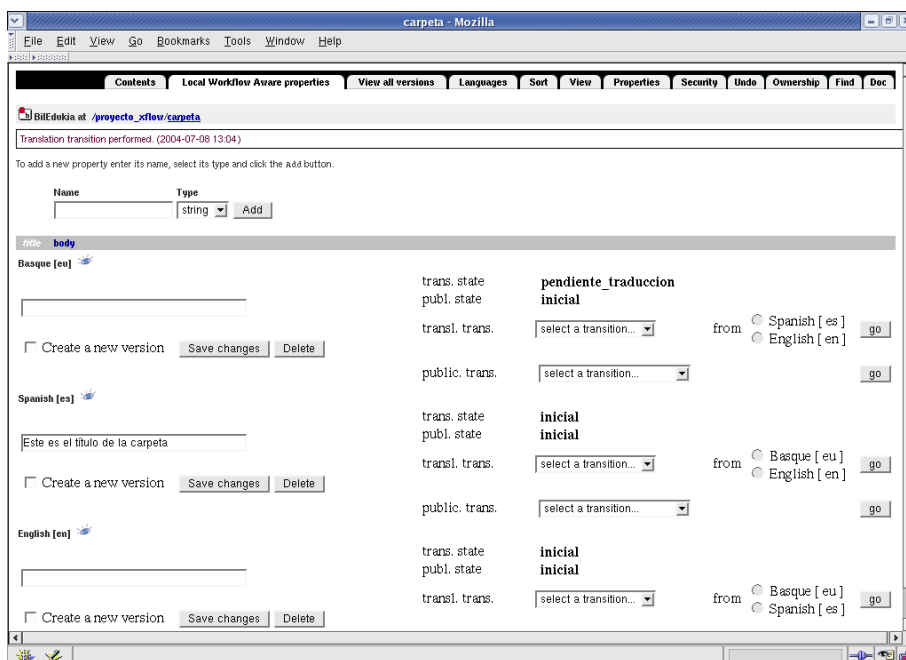


Figura 4.3: Pantalla: Petición de traducción de contenidos (2)

- d. Un botón permitirá exportar en formato TMX la memoria de traducción de dicho documento.
- e. Otro botón permite descargarte en formato XLIFF las traducciones y sus cambios a través de las versiones.
- f. Mediante unos controles se decide que dirección va a tomar el documento dentro del flujo de trabajo. Aquí se muestran los posibles destinos que tiene este documento (a corregir, a publicar, a revisar, etc...) y se decide a donde enviarlo.

4.4.4. El revisor

Determinadas personas tendrán la responsabilidad de controlar el material que se publica. Estos serán los revisores de contenidos. Los revisores de contenidos serán de dos tipos: correctores y moderadores. Los primeros se adecuan más a un perfil lingüístico mientras que los segundos están más relacionados con el mensaje que transmiten los contenidos.

Las pantallas con las que cuenta el revisor son muy similares a las del traductor, sin embargo, el revisor no tiene por qué saber más de un idioma por lo que los contenidos serán presentados en un solo idioma:

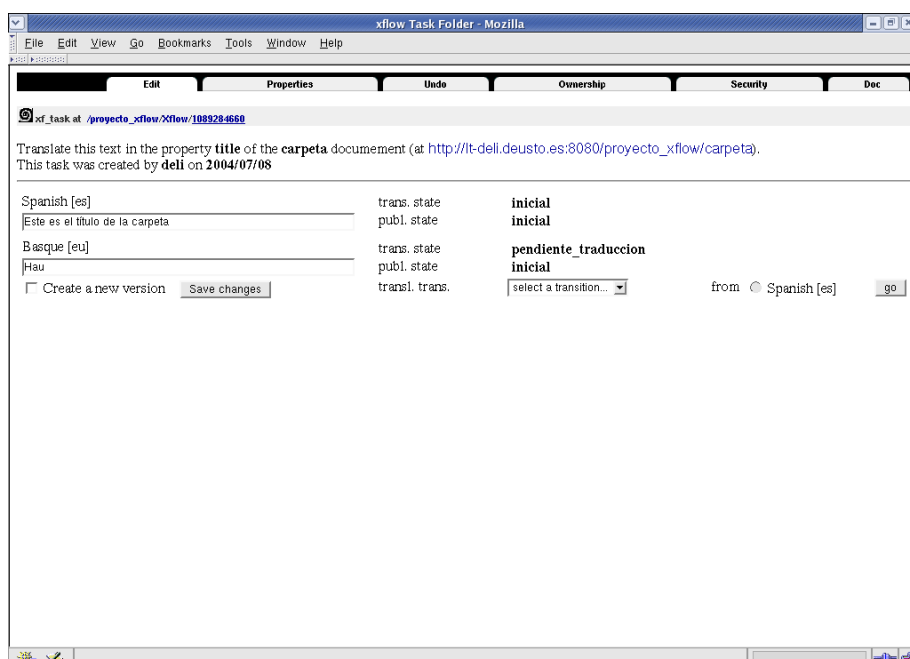


Figura 4.4: Pantalla: Tarea de traducción

- a. Una lista de tareas de corrección pendientes con la siguiente información:
 - I. Quién ha creado la tarea de corrección,
 - II. Fecha de creación,
 - III. Prioridad,
 - IV. Número de palabras,
 - v. Título (primera linea, etc. . .),
 - VI. Tipo de corrección (Ortografía y estilo / por el moderador),
 - VII. Categoría del documento (o si se trata de un mensaje),
 - VIII. Idioma,
 - IX. El estado en que se encuentra este documento en otros idiomas.
- b. Una serie de controles permitirán filtrar el listado de las tareas u ordenarlas en función de algún criterio.

Cuando el traductor selecciona una tarea de traducción, se encontrará con una pantalla (ver figura 4.5) en la que podrá realizar la traducción. Es importante mantener el contexto del texto a traducir por lo que se mostrará el documento completo en el que se encuentre el texto a traducir. En esta pantalla aparecerán varios elementos:

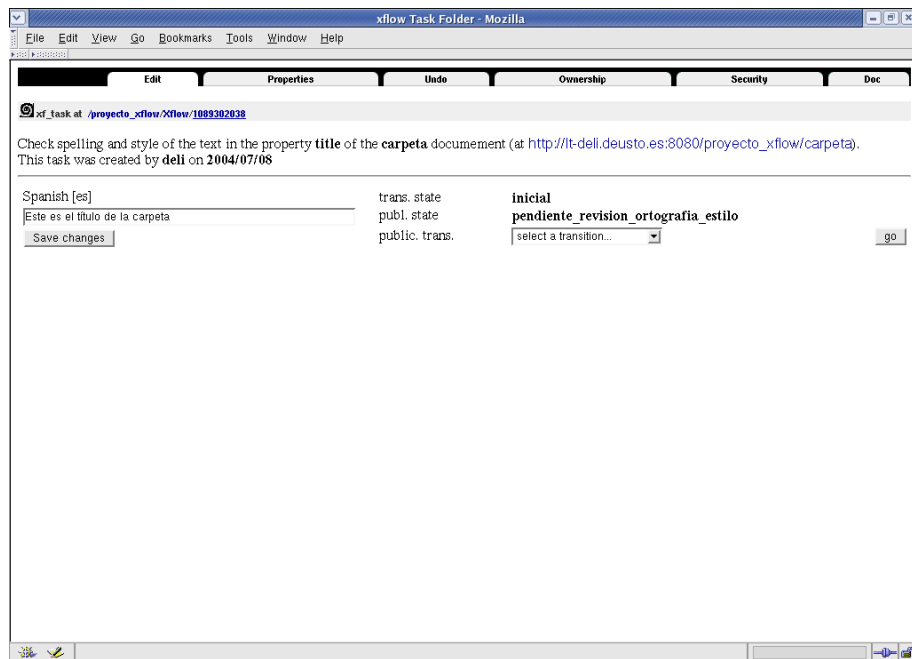


Figura 4.5: Pantalla: Tarea de revisión

- a. Mediante pestañas (tabs) u otros controles se podrán acceder a las distintas versiones de este documento.
- b. El texto del documento en los idiomas origen y destino alineados.
- c. El texto del documento en los idiomas origen y destino en forma de texto corrido.
- d. Un botón permitirá exportar en formato TMX la memoria de traducción de dicho documento.
- e. Otro botón permite descargar en formato XLIFF las traducciones y sus cambios a través de las versiones.
- f. Mediante unos controles se decide que dirección va a tomar el documento dentro del flujo de trabajo. Aquí se muestran los posibles destinos que tiene este documento (a corregir, a publicar, a revisar, etc.) y se decide a donde enviarlo.

4.4.5. Otros

Los usuarios serán los visitantes de un determinado sitio web. No tienen por qué dominar más de un idioma.

En el sistema, una persona cuyas características encajan en un determinado grupo adquiere este rol. A menudo existirán personas que encajen en distintos grupos. En cada interacción con el sistema adquirirán un rol u otro en función de la tarea que estén realizando. Por ejemplo, una persona bilingüe con amplios conocimientos en lenguajes de programación, cuando esté programando en XHTML una nueva pantalla de interfaz de usuario adquirirá el rol de programador, sin embargo cuando entre al sistema para traducir un texto estará bajo el rol de traductor.

Capítulo 5

Fases de un documento multilingüe

Un documento multilingüe estará formado por un conjunto de versiones. Cada una de estas versiones estará compuesta por varios textos, uno en cada idioma. Estos textos pasan por distintos estados (ver figuras 5.1 y 5.2). Los estados por los que puede pasar un documento son de distinta naturaleza. Se podría entender que el documento avanza en dos planos:

1. El plano de la publicación. Dentro de este plano, existen varios estados.
 - a. Inicial.
 - b. Pendiente de revisión (ortografía y estilo). El documento está a la espera de que una persona con conocimientos filológicos revise el documento.
 - c. Pendiente de revisión por el moderador. La persona responsable de los contenidos que se publican debe dar el visto bueno al documento.
 - d. Oculto. El responsable de contenidos pospone la publicación de dicho contenido.
 - e. Visible. El responsable de contenidos ha aceptado el contenido del documento.
 - f. Rechazado por el moderador. El responsable de contenidos rechaza de forma definitiva la publicación de dicho documento.
2. El plano de la traducción. En este otro plano, el documento atraviesa varios estados:
 - a. Inicial.
 - b. Pendiente de traducción. El documento está a la espera de que un traductor lo traduzca.

- c. Pendiente de revisión de la traducción. La traducción deberá ser revisada por otro traductor.
- d. Visible. El documento está correctamente traducido.

5.1. Los estados de un documento

El documento, sus versiones y sus estados pueden organizarse en una especie de matriz en la que se cruzan las versiones y los idiomas del documento.

Estados de un documento.				
	t0. Inicial	t1. Pendiente de traducción	t2. Pendiente de revisión de la traducción	t3. Visible
p0. Inicial				
p1. Pendiente de revisión (ortografía y estilo)				
p2. Pendiente de revisión por el moderador				
p3. Oculto				
p4. Visible				
p5. Rechazado por el moderador				

Figura 5.1: Estados de un documento.

Estructura de un documento.				
	es	eu	en	...
versión 1				
	p5+t3	p5+t1	p5+t3	
versión 2				
	p5+t3	p5+t3	p5+t1	
versión 3				
	p4+t3	p4+t1	p4+t2	
...				

Figura 5.2: Estructura de un documento.

Capítulo 6

El producto X-Flow

X-Flow es un producto Zope cuya misión es agilizar y mejorar la gestión del workflow de una organización entendiendo como tal en nuestro caso el correcto flujo de documentos multilingües desde su incorporación al sistema en su versión monolingüe hasta su publicación web en varias lenguas.

Se identifican los siguientes perfiles de usuario del sistema:

1. programadores: encargados de configurar y mantener el sistema de control de flujo documental.
2. editores de contenidos: responsables de introducir y mantener los contenidos de un sitio web y de escribir los mensajes del interfaz de usuario.
3. traductores: dominan dos o más idiomas y se encargan de traducir tanto los contenidos documentales del sistema como los mensajes de su interfaz.
4. visitantes: no tienen por qué conocer más de un idioma, a ellos está destinado el sitio web.
5. revisores de contenidos: controlan el material publicado, tanto en su corrección lingüística como en el mensaje transmitido.

Para cada uno de estos perfiles se ha definido un interfaz de usuario adaptado a sus necesidades y características concretas. Se tiene por lo tanto en cuenta no sólo el rol del usuario sino las capacidades o competencias concretas de cada persona que desempeña un rol.

Las funcionalidades aportadas por X-Flow difieren para los diferentes tipos de usuario.

El editor de contenidos, dependiendo de si va a editar mensajes cortos para el interfaz del sistema multilingües o a aportar documentos más extensos tendrá unas posibilidades u otras de interacción con el sistema.

Durante la edición de mensajes cortos se puede:

1. acceder a las diferentes versiones del documento
2. introducir el texto del mensaje en una o varias lenguas
3. dirigir el mensaje en el flujo de trabajo
4. aportar propiedades al mensaje
5. descargar el mensaje en formato XLIFF

Durante la edición de documentos de contenidos extensos las opciones serán las anteriores, adaptadas a las características de estos documentos. Además se podrá exportar en formato TMX la memoria de traducción obtenida a partir de la versión multilingüe del documento.

El traductor podrá acceder al listado de tareas pendientes, pudiéndolas filtrar u ordenar por diferentes criterios. Durante el proceso de traducción, el traductor tendrá acceso a:

1. las diferentes versiones del documento
2. los textos alineados
3. herramientas de exportación a TMX y XLIFF
4. controles que le permitan decidir el siguiente destino del documento (corrección , revisión, publicación, etc.)

El revisor dispondrá de las siguientes facilidades:

1. acceso a las tareas pendientes
2. posibilidad de ordenar y filtrar las tareas por diferentes criterios
3. Los documentos le serán presentados en una única lengua.

El programador deberá poder conocer el estado de los diferentes documentos y tener acceso a sus diferentes versiones.

Un documento multilingüe avanza por diferentes versiones. Cada una de estas versiones está compuesta por varios textos, uno en cada lengua. A su vez, cada uno de estos textos pasan por una serie de estados, al largo en dos dimensiones (publicación y traducción) como se describe en el capítulo 5.

X-Flow adopta las características de los productos Localizer y Flux aprovechando la flexibilidad que el paradigma de la programación orientada a objetos ofrece. En algunas ocasiones ha sido necesario reimplementar alguna de las funciones de Localizer y Flux para completar con éxito los objetivos del prototipo de X-Flow. El diagrama 6.1 clarifica el diseño del prototipo, donde se pueden ver las relaciones entre las clases de xflow y las de otros módulos.

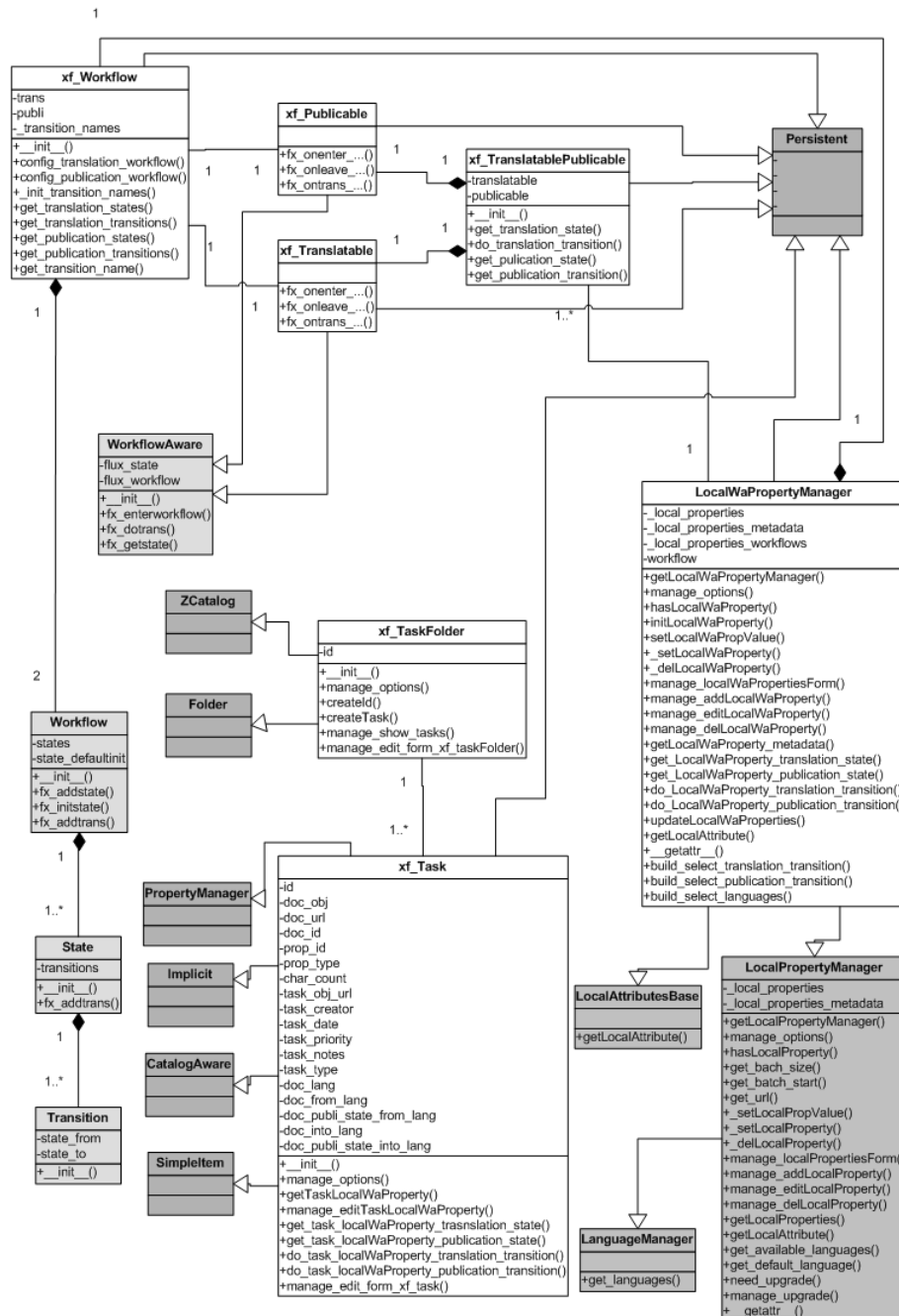


Figura 6.1: Diagrama de classes

Capítulo 7

Aportación de X-Flow

Los objetivos alcanzados podrían resumirse en estos:

1. Aplicación en un contexto industrial- comercial de la investigación realizada por el equipo de proyecto relativa a la gestión de corpus paralelos multilingües y a los filtros de conversión de formatos estándar para el almacenamiento e intercambio de información multilingüe (TMX y XLIFF)
2. Reutilización y aplicación de recursos lingüísticos y herramientas desarrolladas en proyectos anteriores.
3. Enriquecimiento mutuo entre universidad y empresa en el área de la ingeniería lingüística mediante el intercambio de conocimientos y aplicaciones.
4. Diseño de un prototipo de producto Zope que permitirá la agilización de la actualización de contenidos tanto en la administración como en el mantenimiento de portales multilingües
5. Difusión de la actividad realizada mediante la presentación de 8 contribuciones en congresos nacionales e internacionales y revistas científicas.
6. Transferencia tecnológica a la empresa que actualiza su know-how aplicable en la prestación de servicios y el desarrollo de nuevos productos ofrecidos a la sociedad.

Bibliografía

- David Hollingsworth. 1995. Workflow Management Coalition: The workflow Reference Model. <http://www.wfmc.org/standards/docs/tc003v11.pdf>.
- Elmagarmid, Ahmed y Weimin DU. 1998. Workflow management: State of the art versus state of the products. En *DOĞAÇ, Asuman; et al., editor, Workflow Management Systems and Interoperability*, volumen 164 de *Nato ASI Series. Series F: Computer and Systems Sciences*. Springer-Verlag Berlin Heidelberg, New York, páginas 1–17.
- Fisher, Layna. 2003. *Workflow Handbook 2003*. Future Strategies Inc.
- Free Software Foundation. 2002. GNU Project/Gettext. <http://www.gnu.org/software/gettext>.
- Josu Gómez. 2001. Una guía al TMX. <http://sirio.deusto.es/abaitua/deli/xtrabi-e341.htm>.
- Karasek, Milan. 2001. Xliff, a new standard in localization. *Moravia IT News*.
- Localization Industry Standars Association (LISA). 2002. LISA/OCAR-TMX-Translation Memory eXchange eStandard. <http://www.lisa.org/tmx/>.
- OASIS XLIFF TC. 2003. OASIS XML Localisation Interchange File Format TC. <http://www.oasis-open.org/committees/xliff>.
- Palomar, Juaán David Ibañez. 2002. Localizer. *V Congreso HispaLinux*.
- Palomar, Juán David Ibáñez. 2003. Localizer home. <http://www.localizer.org/>.
- Savourel, Yves. 2001. *XML Internazionalization and Localization*. Sams Publishing.
- Simon, Alan R. y William Marion. 1996. Workflow technology. En *Workflow computing: Workflow, Groupware and Messaging*. McGraw-Hill, New York, páginas 47–61.
- Zope Community. 2003. Zope Home Site. <http://www.zope.org>.