

# Informe de Investigación

Fernando Quintana Hernández

# Índice general

<b>1. Introducción</b>	<b>3</b>
<b>2. Objeto</b>	<b>4</b>
<b>3. Método</b>	<b>5</b>
3.1. Documentación . . . . .	5
3.2. Desarrollo de la solución . . . . .	6
3.3. Divulgación . . . . .	6
<b>4. Herramientas</b>	<b>7</b>
<b>5. Resultados</b>	<b>9</b>
<b>6. Conclusiones</b>	<b>10</b>

# Capítulo 1

## Introducción

Una vez superado el periodo de docencia, cada doctorando del programa de Ciencia de la computación debe conseguir 12 créditos mediante trabajos de investigación tutelados. Durante el curso académico 2003-2004 han sido dos los trabajos de investigación que han ocupado mi tiempo: “El sistema de gestión de traducciones SareBi” y “Xflow: Gestión de flujo de contenidos multilingües sobre XLIFF y TMX”.

Con este informe pretendo resumir el objeto de la realización de ambos títulos así como describir los métodos y herramientas que he empleado durante su desarrollo y reflexionar sobre los resultados obtenidos y las conclusiones a las que los presentes trabajos me han llevado.

Durante los últimos dos años, he tenido la suerte de trabajar como investigador ayudante en el seno del grupo DELi, formado por profesores de las facultades de Letras y ESIDE con intereses comunes en las áreas de edición digital e ingeniería lingüística. Esta colaboración me ha brindado una oportunidad inmejorable para dar mis primeros pasos en el mundo de ciencia de la mano de experimentados investigadores como son JosuKa Díaz, Joseba Abaitua o mi tutora Inés Jacob.

A lo largo de estos años hemos trabajado en varios proyectos de distinta índole, siempre en torno a la problemática de la edición digital y el multilingüismo. Los trabajos que he presentado tratan de resumir el esfuerzo de investigación realizado durante el desarrollo de dos de ellos.

El siguiente capítulo atiende a la motivación que me ha llevado a realizar estos trabajos y cuales han sido las circunstancias de partida del periodo de investigación. El segundo capítulo es un resumen de la metodología utilizada durante el desarrollo de las investigaciones. Un tercer capítulo recopila las herramientas, técnicas y otros recursos de los que he hecho uso. Los resultados obtenidos durante el periodo de investigación aparecen en el cuarto capítulo, donde se hace un somero repaso de los objetivos y su grado de resolución. El último capítulo recoge una valoración personal a modo de conclusión.

## Capítulo 2

# Objeto

Además de permitirme optar al Certificado-Diploma de Estudios Avanzados, los trabajos debían servirme como un ensayo metodológico de las investigaciones que he de llevar a cabo a lo largo del próximo año académico, durante la realización de mi tesis doctoral.

En esta línea, el escenario de trabajo ha sido real, por lo que la experiencia iba a darme a conocer la dinámica general de un grupo de investigación e iba a permitirme observar en primera persona las pautas de conducta en distintas situaciones en las que se iban a desarrollar las tareas de investigación (reuniones con usuarios y empresas, organización y asistencia a congresos, colaboraciones con otros grupos de investigación, etc.)

Una de las primeras batallas que tiene que librar un doctorando es la búsqueda del tema de su tesis. Esta lid es mucho más fácil si, como es mi caso, cuenta con la ayuda de un grupo consolidado, con experiencia y tradición investigadora en un determinado campo del conocimiento.

En ese sentido, durante el desarrollo de estos trabajos, y en general de los proyectos de investigación en los que he participado, gran parte del esfuerzo realizado ha sido dirigido hacia el objetivo de seleccionar el tema concreto sobre el que versará mi tesis doctoral. De esta forma, para la realización de cada trabajo de investigación, se ha llevado a cabo un amplio estudio del estado de la cuestión en los campos en los que se enmarcan ambos trabajos.

## Capítulo 3

# Método

### 3.1. Documentación

Todo trabajo de investigación comienza con un arduo trabajo de documentación. Una vez decidida la problemática que se va a abordar, es necesario conocer a fondo todos los trabajos que anteriormente se han realizado sobre el tema en cuestión.

Para llevar a cabo esta labor, han sido numerosas las fuentes consultadas y de diversa naturaleza. Los fondos de la biblioteca universitaria y la documentación en formato electrónico publicada en Internet han sido los principales recursos explotados en las tareas de documentación. La búsqueda de información en la biblioteca, concierne dos tipos de tareas:

- a. En primer lugar, es posible efectuar consultas al catálogo electrónico, ya sea vía web<sup>1</sup> desde cualquier ordenador conectado a Internet o mediante los terminales dispuestos específicamente para ello en la propia biblioteca.
- b. Por otro lado, los investigadores tienen acceso físico a los fondos bibliotecarios de la universidad, es decir, es posible ojear distintos obras hasta encontrar la información deseada. Además del depósito principal de libros, existen otras salas donde se pueden encontrar documentos de otros géneros, como es el caso de la sala de publicaciones periódicas. En esta sala se pueden consultar los últimos números de la gran mayoría de la prensa especializada que se publica, lo cual supone un valioso recurso para conocer el estado actual del objeto de un estudio.

Pese a todo, la principal fuente de información consultada no ha sido la bibliografía escrita. A parte de los conocidos buscadores de propósito general, es importante destacar la utilidad de los distintos repositorios de

---

<sup>1</sup><http://www.biblioteca.deusto.es>

documentación científica como el buscador SiteSeer<sup>2</sup>, o el portal de la ACM<sup>3</sup>.

Estos sitios web permiten realizar búsquedas de forma algo más sofisticada que los catálogos electrónicos de la biblioteca universitaria, por eso es importante visitar estas webs antes de ratrear la biblioteca.

### **3.2. Desarrollo de la solución**

Una vez recopilada y estudiada toda la documentación, y tras haber planteado la solución al problema que es objeto de estudio, comienza la fase de desarrollo de la misma. Antes de comenzar a implementar la solución es necesario realizar un análisis de requisitos y un diseño de la aplicación. Solo después de estas tareas, cuando las características de la solución están bien definidas, comienza la fase de implementación.

### **3.3. Divulgación**

Un trabajo científico debe ser sometido a la opinión crítica de la comunidad especializada. Esta labor de divulgación es especialmente importante a la hora de contrastar la hipótesis planteada, escrutando el nivel de aceptación que tiene la solución propuesta entre la comunidad científica. La divulgación de los trabajos realizados presenta una buena oportunidad al equipo investigador para que evaluadores externos evalúen sus trabajos y ofrezcan una visión crítica de los resultados.

---

<sup>2</sup><http://citeseer.ist.psu.edu/>

<sup>3</sup><http://portal.acm.org/dl.cfm>

## Capítulo 4

# Herramientas

Como ya he mencionado en el capítulo 3, las herramientas más utilizadas han sido las de búsqueda y extracción de información. Durante el trabajo de documentación han sido numerosos los títulos de libros y artículos, manuales y tutoriales que he utilizado. Los documentos más destacados aparecen referenciados en las distintas bibliografías de los trabajos.

La solución ha sido implementada sobre la plataforma Zope. Esta plataforma está desarrollada casi en su totalidad en Python y su arquitectura incluye una base de datos orientada a objetos ZODB (Zope Object Data Base) como almacén de información. Todos los elementos de una aplicación Zope se elaboran como objetos Python de distintas clases para su publicación en la web (en cualquiera de los formatos HTML, XHTML, XML, u otro).

Para el desarrollo de la documentación elaborada a lo largo de este periodo de investigación, he utilizado diversos procesadores de texto y herramientas de edición digital. En general, tanto a la hora de generar documentación, como a la hora de implementar aplicaciones, el grupo DELi ha optado siempre por soluciones en forma de software libre como OpenOffice, Xemacs o L<sup>A</sup>T<sub>E</sub>X.

T<sub>E</sub>X es un sistema de composición tipográfica de textos de alta calidad y refinamiento, y con funciones avanzadas de automatización. T<sub>E</sub>X en sí mismo es un completo lenguaje de programación que permite el marcado lógico de documentos de forma que el formato se pueda generar automáticamente. L<sup>A</sup>T<sub>E</sub>X es un paquete de macros para T<sub>E</sub>X, originalmente escrito por Leslie Lamport para proporcionar un sistema de procesamiento de documentos más simple de uso que T<sub>E</sub>X, pero con toda su potencia.<sup>1</sup>

Existen distintos medios para divulgar las soluciones propuestas o publicar los resultados de una investigación, por ejemplo leyendo una comunicación en algún congreso o escribiendo un artículo para alguna revista. A lo largo del curso, he tenido la oportunidad de tomar partida en ambas

---

<sup>1</sup><http://www.cervantex.org>

modalidades. Algunos de estas publicaciones y comunicaciones aparecen en los anexos de los trabajos de investigación entregados.

La lingüística computacional es una área multidisciplinar, por lo que el público ante el cual es necesario divulgar los resultados es especialmente plural. En este sentido, el esfuerzo de publicación debe ser más intenso para llegar a todos los grupos de expertos a los que les puede interesar la investigación realizada.

Por otra parte, el sitio web del grupo de investigación<sup>2</sup> también ha servido de canal de difusión de los avances en los proyectos realizados.

---

<sup>2</sup><http://www.deli.deusto.es>



## Capítulo 5

# Resultados

Estos dos años desempeñando labores de investigación han resultado de gran provecho. En términos generales, un estudiante de ingeniería informática no conoce apenas el mundo de la investigación y la ciencia cuando finaliza la carrera. La toma de contacto con este mundo se me ha presentado en este periodo de investigación. En este sentido, el ensayo metodológico que iba a resultar este curso de doctorado ha resultado positivo.

La participación en la disciplina cotidiana de un grupo de investigación consolidado me ha permitido adquirir una visión clara del panorama de la investigación y ciencia, sobre todo en el área de la lingüística computacional. En esta línea, he podido conocer la dinámica de publicación de artículos y de asistencia y organización de congresos así como intercambiar impresiones con numerosas figuras del área.

Del mismo modo, he podido presenciar los prolegómenos de una defensa de una tesis, o la evolución de la presentación escrita de la misma. Estas experiencias me permitirán actuar con más seguridad cuando de mi tesis se trate.

Paralelamente he conseguido seleccionar el tema de tesis, lo cual supone un gran paso en firme hacia la consecución del grado de doctor. He adjuntado un resumen del mismo en el anexo II de este informe. La propuesta fue bien acogida por la comunidad científica cuando fue presentada en el seminario<sup>1</sup> sobre procesamiento de lenguaje natural que tubo lugar el pasado febrero en Hondarribia.

---

<sup>1</sup><http://ixa.si.ehu.es/Ixa/Bestelakoak/ws/programa>

## Capítulo 6

# Conclusiones

La experiencia de trabajar durante dos años como ayudante de investigación me permite decir que fue un acierto ingresar en el grupo DELi y no puedo sino agradecer a sus miembros el haberme aceptado y tratado de la forma que lo han hecho tanto en el plano profesional como en el personal y especialmente a mi tutora Inés por orientarme con sutil firmeza a lo largo de estos dos últimos años.

Me considero en una posición privilegiada respecto a aquellos aspirantes a doctor que no han podido integrarse en un grupo de investigación y sin duda, siento que cuento con posibilidades reales de finalizar la tesis doctoral en los próximos años. Sin embargo, en este punto, queda en manos de la Comisión de Doctorado la valoración de los trabajos presentados y con ello la posibilidad de optar al Certificado-Diploma de Estudios Avanzados.