

Capítulo 1

Introducción

El formato PDF (*Portable Document Format*) es en la actualidad el estándar de facto en todo el mundo para el intercambio de documentos tanto a nivel empresarial y gubernamental como de usuario. Conforman hoy en día el medio por excelencia para la transmisión de publicidad, formularios, manuales, proyectos, etc. como pone de manifiesto el hecho de que podamos encontrar en la red más de 200 millones de documentos PDF [18]. Su éxito viene dado principalmente por su carácter multiplataforma. PDF puede visualizarse e imprimirse de forma indistinta en sistemas Macintosh, Microsoft Windows, Unix y otros dispositivos móviles.

Desafortunadamente, la mayoría del software capaz de generar, interpretar y modificar ficheros PDF es de carácter privativo y los usuarios se ven obligados a pagar costosas licencias para poder adaptarse a una realidad aplastante: el dominio de PDF en el ámbito de los documentos electrónicos imprimibles. Según *Free Software Foundation*, con la inminente estandarización de PDF 1.7 en la ISO 32000, que extenderá aún más si cabe su uso, el acceso libre a la tecnología PDF resulta una cuestión urgente [43].

Así, este proyecto pretende cooperar en la misión de crear herramientas de software libre entorno a *Portable Document Format*. En concreto, se proporciona una herramienta libre de *modificación* de PDF que es capaz de solucionar uno de los problemas más frecuentes de los usuarios en la manipulación de estos ficheros, como es la no incrustación de las fuentes dentro de los PDF¹. Si una fuente no se encuentra incrustada, el visualizador o impresora que gestiona el documento tiene dos opciones. La primera es adaptar una de sus fuentes disponibles para sustituir de forma aproximada a la solicitada. La segunda opción es la no representación del texto. Es decir, este problema añade una falta de portabilidad y de mantenimiento de la integridad de la información relevante para un tipo de documento que se caracteriza por su autonomía e independencia de la plataforma.

La solución que se aporta pretende ser de gran trascendencia en la transmisión de documentos electrónicos a conferencias y otros eventos, así como en el almacenamiento en librerías

¹La incrustación de fuentes suele ser omitida bien por desconocimiento del creador del documento, por la incapacidad del *distiller* (programa de creación de PDF) o por la indisponibilidad de las fuentes en el sistema.

digitales de acceso a información técnica de gran calidad. De hecho, de la especial atención al formato de los documentos PDF que presta la asociación IEEE (Institute of Electrical and Electronics Engineers, Inc.) surge la idea inicial de este proyecto. IEEE [48] delega en IEEE Xplore® para la gestión de sus revistas, transacciones, cartas, estándares, artículos de conferencias y publicaciones de IET (Institution of Engineering and Technology). IEEE Content Engineering define los requisitos para la aceptación de documentos PDF o cumplimiento de las normas Xplore mediante PDF Specification for IEEE Xplore® [50]. Los criterios de admisión se definen en base a propiedades que determinan su calidad y capacidad de visualización en distintas plataformas y sistemas de todo el mundo. Entre las exigencias destaca la incrustación de **TODAS** las fuentes que se emplean [49]. Así, Xplore presenta en su última catálogo de normas (a mayo de 2007) la siguiente advertencia al respecto de este problema [51]:

“Once IEEE receives a PDF file without embedded fonts *nothing can be done*. Using these fonts may *delay posting* the issue or possibly cause the paper *not to be included* in IEEE Xplore®”

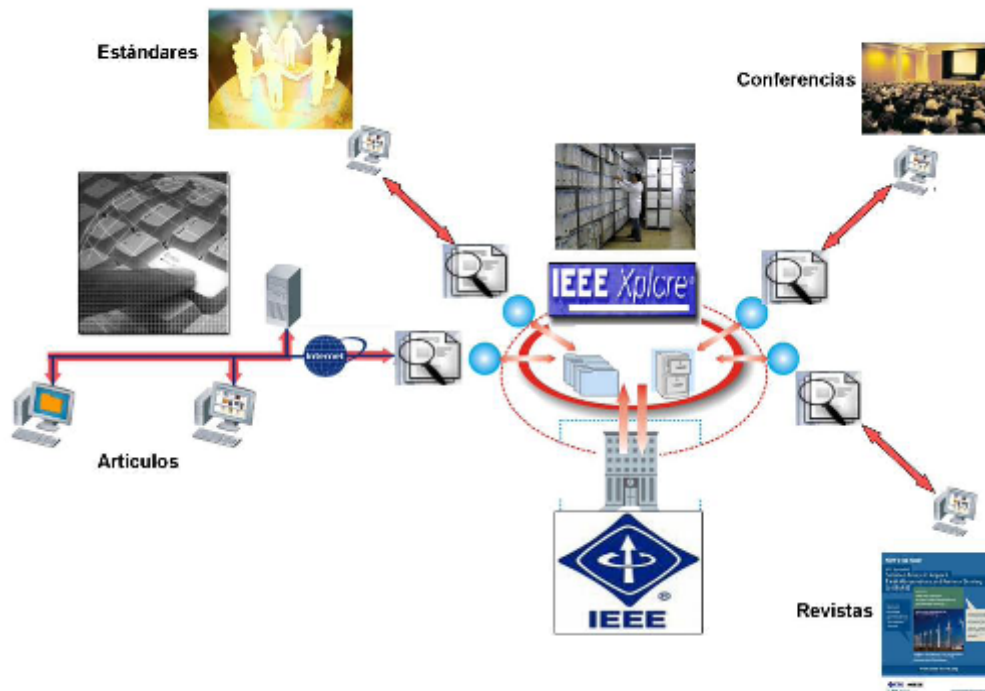


Figura 1.1: Transmisión de documentos PDF y verificación de requisitos para aceptación en la librería electrónica de IEEE, IEEE Xplore.

El problema de la mayoría de los *distillers* es que por defecto **NO** incrustan (o simplemente no tienen la capacidad para incrustar) las 14 fuentes base o *base 14 fonts* ni las demás fuentes

disponibles en el sistema, a no ser que se específicamente se le indique (en caso de que exista esa posibilidad). Además, todas las fuentes que se desean incrustar deben encontrarse en el sistema en el que se crea el PDF. Dado que existen en el mercado gran cantidad de soluciones para la creación de documentos electrónicos *Portable Document Format*, es prácticamente imposible proporcionar las directrices de preparación para cada sistema y *distiller* de forma que se ajuste a las especificaciones IEEE en cuanto a la gestión de fuentes. Es por esto que resulta necesario que la solución a este problema sea independiente del autor.

La aplicación que se desarrolla en este Proyecto Fin de Carrera intenta solucionar este problema proporcionando una herramienta *open source* que incruste las fuentes que se emplean en un documento PDF, que no necesite disponer del documento fuente original, padre del PDF generado y consiga así dar un paso adelante en el acceso libre a *Portable Document Format* y en el procesamiento automático de artículos remitidos a conferencias y demás disertaciones para el cumplimiento de los requisitos de calidad.

1.1. Enfoque

Las herramientas para la visualización y creación de documentos PDF tanto en software libre como propietario son abundantes, sin embargo, la capacidad para la modificación de los documentos resulta una cualidad escasa de las aplicaciones y librerías entorno a PDF, especialmente en software no privativo.

En este proyecto se persigue el diseño y creación de una herramienta de análisis e incrustación de fuentes en lenguaje Perl [53] empleando el módulo PDF::API2 [28] como base de la aplicación. Con la incrustación de fuentes se garantiza que el PDF use las mismas fuentes que el documento original, con independencia de dónde se visualice o imprima el PDF o de las fuentes que se tengan instaladas en ese equipo.

A la aplicación se accederá de dos formas. Mediante línea de comandos y además, para favorecer la usabilidad, el acceso a la lógica central y gestión de recursos, se genera una interfaz gráfica de usuario también en Perl, con el soporte principal del módulo Tk. Se pretende en todo caso favorecer el mantenimiento de la herramienta software, siguiendo para ello un diseño multicapa.

1.2. Estructura del trabajo

El documento se divide en ocho capítulos y cuatro anexos:

- **Capítulo 1**, Introducción. Introducción, enfoque del proyecto y estructura del documento.
- **Capítulo 2**, Objetivos del proyecto. Enumeración y descripción de cada uno de los objetivos propuestos.

- **Capítulo 3**, Antecedentes I: Estado del arte. En este capítulo se sitúa a *Portable Document Format* (PDF) dentro del mundo de los documentos electrónicos y se presentan sus características principales. Además, se realiza un estudio sobre las diferentes tecnologías de fuentes que soporta PDF y se comenta cual es la situación actual de las herramientas y aplicaciones tanto en software libre como propietario entorno a dicho formato, y en especial las más competitivas entorno a la gestión de fuentes.
- **Capítulo 4**, Antecedentes II: Arquitectura básica de PDF. Breve introducción a los conceptos fundamentales de estructura y organización de los documentos PDF.
- **Capítulo 5**, Antecedentes III: Gestión de fuentes en PDF. Análisis de las capacidades, estructuras y entidades básicas de PDF para reproducción de textos y tratamiento de fuentes.
- **Capítulo 6**, Métodos y fases del trabajo. En este capítulo se describen los diferentes elementos que componen el sistema, métodos empleados y fases del ciclo de vida del software que se han llevado acabo para la realización del proyecto.
- **Capítulo 7**, Resultados. Resultados obtenidos tras la realización del proyecto y un análisis de las posibles formas de licencia con que se podría distribuir la aplicación.
- **Capítulo 8**, Conclusiones y propuestas. Cuales han sido las conclusiones que se han obtenido, y algunas propuestas que pueden mejorar el sistema actual.
- **Anexos:**
 - **Anexo A:** Manual del usuario.
 - **Anexo B:** Estructura del código fuente.
 - **Anexo C:** Instalación.
 - **Anexo D:** Código fuente de la aplicación.