

Proyecto GIDIN - Gestor Inteligente de Documentación Incoherente

Autores:

Sergio Martín
Juan Peire
Manuel Castro

Dep. de Ingeniería Eléctrica Electrónica y de Control
Universidad Nacional de Educación a Distancia



Financiado por:
Terasoft Consultores S.L.

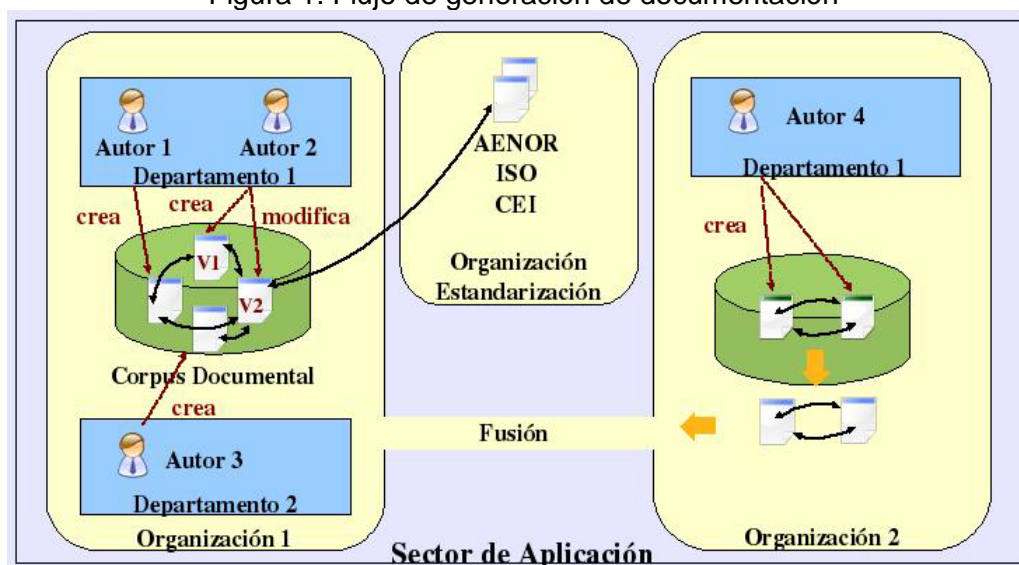
Madrid, 10 de Febrero de 2011

1. Introducción

La documentación, tanto en papel como en formato electrónico, es un elemento fundamental de la sociedad de la información. Es el modo más usado para almacenar, guardar e intercambiar información en la mayoría de las actividades desarrolladas actualmente en el contexto humano, por lo que la información y el conocimiento que contienen debe ser correcto y claro, evitando así cualquier posibilidad de confusión y contradicción. Pero este objetivo no es trivial y depende de múltiples factores.

Existen en la actualidad ciertos sectores, públicos y privados, en los que se maneja gran cantidad de documentación para llevar a cabo su función empresarial u organizativa. Debido a esta actividad a lo largo de los años, y a la posible fusión con otras empresas o departamentos, esta documentación sufre cambios en su elaboración, además de crecer en volumen y versiones. De esta manera resulta difícil encontrar un conjunto de documentación que se haya originado desde una única fuente documental, y lo más habitual es que las organizaciones dispongan de una documentación generada por múltiples autores y desde múltiples fuentes documentales (áreas, departamentos o sucursales).

Figura 1: Flujo de generación de documentación



Es difícil encontrar organismos que gestionen todo este movimiento de documentación de una manera centralizada y formal, por lo que suele ser habitual que dichas entidades cuenten con un conjunto de documentación heterogénea, que no ha seguido una formulación única en su generación, y con múltiples estructuras (con información estructurada, no estructurada o semi estructurada). Este hecho genera ciertos problemas, que se centran en la aparición de incoherencias o “inconsistencias” en dicha documentación. Esta situación suele traer consigo una serie de problemas derivados que se generalizan en la incapacidad de realizar una gestión coherente de la documentación.

Las consecuencias que se extraen de la existencia de inconsistencias, y de la incapacidad de realizar una gestión documental coherente, se generalizan en la aparición de problemas no solo técnicos, sino también en problemas legales, sociales y económicos, al disponer de una documentación que puede provocar confusiones, errores o perjuicios a las entidades usuarias de la misma.

La aparición de estos problemas, que no solo afectan a la entidad u organización generadora de la documentación, sino que también puede llegar a provocar repercusiones sociales, parecen justificar el interés en estudiar las técnicas que la ciencia y la tecnología ponen a disposición del hombre para la detección de estas inconsistencias, su eliminación y el establecimiento de alguna metodología que evite su repetición.

De esta manera, el presente proyecto centra su actividad en conseguir dos objetivos distintos pero unidos en la consecución de un único fin, la obtención de un **conjunto de documentación coherente**. Para ello los objetivos generales perseguidos son dos:

1. Detectar los problemas existentes en la documentación actual que se maneja en las organizaciones.
2. Evitar la repetición de estos problemas detectados en la generación de nueva documentación.

Con esta finalidad se ha planteado el proyecto “Gestor Inteligente de Documentación Incoherente”, de aplicación a tres sectores de actividad:

- Sector eléctrico: a través del estudio de la documentación de la compañía Iberdrola Distribución Eléctrica, S.A.
- Sector ferroviario: a través del estudio de la documentación de la compañía ADIF.
- Sector científico: a través del estudio de la documentación del Real Jardín Botánico de Madrid.

Desde Iberdrola se ha abordado este problema formando un equipo de trabajo multidisciplinar, con la unión de técnicos especialistas procedentes de diferentes empresas, centros y áreas tecnológicas.

La ejecución del proyecto se planteó a través de diferentes fases de ejecución y tareas.

- FASE 1: Análisis del corpus documental.
- FASE 2: Diseño y desarrollo del modelo de datos.

- FASE 3: Análisis de casos de uso.

De forma general, la documentación tratada en este proyecto cumple con las siguientes características:

- Se trata de corpus de documentación formados por una **gran cantidad de documentos**, es decir, se trabaja con un volumen muy grande de documentación.
- Los documentos presentan **relaciones cruzadas**, esto es, contienen referencias a otros documentos, o bien hacen referencia a información contenida en otros documentos o dentro del mismo documento pero en otro punto del mismo.
- En muchas ocasiones, los documentos a los que se hace referencia desde la documentación contienen normativa, protocolos, procedimientos, etc., estandarizados o generados en el propio dominio.
- En los documentos se emplea una **terminología dependiente del campo de aplicación**, y que además debe ser coherente con el contexto.
- Los documentos suelen contener gran cantidad de **tablas, imágenes, figuras, anexos**, e incluso modificaciones del documento original (que pueden aparecer también como anexos del mismo) en el caso de los sectores eléctrico y ferroviario.
- Pueden ser documentos **no estructurados** (sin una estructura definida), **estructurados** (con una estructura definida y conocida) o **semi estructurados** (parte de su estructura está definida, mientras que otras partes son en texto libre).
- Normalmente la documentación ha sido generada por **múltiples autores** y desde múltiples fuentes documentales (departamentos, empresas, entidades, etc.).
- Como consecuencia de lo anterior, la documentación suele presentar **múltiples formatos y estilos**, al no existir una plantilla única de generación compartida por las múltiples fuentes de origen, ni una metodología única de generación de la misma.
- Los documentos están en continua evolución debido a la incorporación de nuevos documentos (nuevas normas, leyes, procedimientos de aplicación a la misma), modificaciones, eliminación de otros documentos obsoletos, etc. que provocan la existencia de **diversas versiones**, principalmente detectado en los sectores eléctrico y ferroviario.
- En ocasiones, la documentación ha sufrido la **fusión o el cambio de propiedad**, por lo que se ha tenido que adaptar a las premisas de su nuevo contexto. Este hecho se produce, por ejemplo, cuando una compañía se fusiona con otra, de forma que es necesario adaptar la documentación manejada por ambas partes con el fin de generar una nueva documentación de aplicación conjunta.