

Generación de Documentos basada en Líneas de Producto

M^a Carmen Penadés¹, José H. Canós¹, Sergio Camarasa¹, Marcos R.S. Borges²,
Adriana S. Vivacqua²,

¹ ISSI - DSIC, Universitat Politècnica de Valencia,
46022 – València, Spain
{mpenades, jhcanos}@dsic.upv.es; sercabom@ei.upv.es

² PPGI-IM/UFRJ, Universidade Federal do Rio de Janeiro,
20010-974 Rio de Janeiro, Brazil
mborges@nce.ufrj.br; avivacqua@dcc.ufrj.br

Resumen. La generación de documentos con contenido variable es cada vez más relevante en muchos dominios, como la administración electrónica o el comercio electrónico, imponiéndose la filosofía de “1 persona-1 documento” vs “1 documento-N personas”. En este artículo presentamos una propuesta metodológica para especificar la variabilidad de una familia de documentos en términos de características, que son transformadas en elementos de contenido mediante editores personalizados que se generan siguiendo una aproximación de Líneas de Producto Software. La propuesta, denominada Líneas de Producto de Documento (*Document Product Line*), se construye sobre un metamodelo de características de documento, que guiará el posterior proceso de generación. Actualmente se está trabajando en un marco de soporte y en su aplicación a la generación de documentos en el dominio de la Gestión de Emergencias.

Keywords: Generación de documentos, Gestión de la Variabilidad, Líneas de Producto Software.

1 Motivación

La generación de documentos personalizados es un aspecto relevante en dominios como la administración electrónica o el comercio electrónico. El contenido de los documentos varía en función de la organización o ciudadano al que van dirigidos, y además, en numerosas ocasiones su estructura está regulada por leyes. Por otra parte, el propio documento se puede generar en diferentes formatos en función de la plataforma tecnológica destino. Se impone el paradigma “1 documento –1 persona” frente a “1 documento–N personas”, es decir, generar documentos gubernamentales, manuales, folletos de propaganda, etc. totalmente personalizados, cuyos contenidos varían en función del destinatario. En este contexto, no han faltado en los últimos años autores que proponen soluciones a la generación de documentos con contenido variable, siendo la mayoría de ellas soluciones tecnológicas basadas en XML [1][2].

Por otra parte, la reutilización de contenido es otro aspecto importante, encontrando en la literatura soluciones basadas en bloques de contenidos reutilizables, como en [3]. También están basadas mayoritariamente en XML, requiriéndose un alto nivel de conocimiento de la tecnología asociada para poder obtener resultados satisfactorios.

En este artículo presentamos un método para la generación de familias de documentos que aborda el problema desde una nueva perspectiva, elevando el nivel de abstracción en el que el ingeniero del dominio especifica las características que comparten y las que difieren el conjunto de documentos que componen la familia. Nuestra propuesta metodológica está basada en los principios, técnicas y tecnología asociada a las Líneas de Producto Software [4][5], por lo se denomina Líneas de Producto de Documento (*Document Product Line*, DPL). DPL especializa el marco definido por las Líneas de Producto Software para el dominio de la generación de documentos, siendo la principal aportación la definición de un metamodelo de características de documento para especificar la variabilidad dentro de una familia. Dicho metamodelo guía el proceso de generación de documentos. La selección de características de documento a incluir en la línea determinará el editor personalizado que se genera, con el que completaremos los datos variables, si existen, en el documento final. Las características de reutilización y personalización implícitas en la ingeniería de Líneas de Producto Software son también proporcionadas por DPL. Se está trabajando en el marco de soporte a la metodología y en su aplicación a la generación de Planes de Emergencia para las organizaciones.

El resto del artículo está organizado como sigue. En la sección 2 se resumen algunos de los trabajos relacionados. En la sección 3 se presenta nuestra propuesta, Líneas de Producto de Documento, que se ilustra brevemente en la sección 4 con un caso de estudio. Para finalizar, las conclusiones y líneas de trabajo abiertas.

2 Trabajos Relacionados

Dentro de la Ingeniería de Documentos, uno de los temas más activos en los últimos años es la generación de documentos con contenido variable. Como ya hemos comentado, gran parte las propuestas son soluciones tecnológicas: documentos XML manipulados con las tecnologías asociadas. Destacamos la propuesta de Lumley *et al.*, que define un lenguaje de descripción de documentos separando los datos, de la estructura lógica, de su presentación [1]. Otra propuesta pero con herramientas de edición basada en la Web es la de Quint [2]. También existen sistemas comerciales con muy buenas prestaciones para el diseño de documentos personalizados. Sin embargo, ninguna de ellas proporciona una clara y completa guía metodológica que permita especificar explícitamente la variabilidad, ni disponen de un proceso de generación del documento automatizado; están más orientadas a la presentación final.

Otros aspectos como la reutilización de contenidos son también especialmente relevantes, siendo el estándar DITA [3] uno de los más conocidos y utilizados. Define una arquitectura para crear y gestionar fragmentos de contenido reutilizable definidos en XML, a los que denomina *tópicos*. Los *mapas* agregan un conjunto de tópicos para crear un documento en distintos formatos (pdf, HTML, etc.) mediante herramientas de publicación asociadas como DITAOpenToolKit (dita-ot.sourceforge.net/). Tanto

en DITA como en otras propuestas similares, la reutilización de contenidos es posible, pero depende totalmente de la habilidad del ingeniero en definir y combinar dichos fragmentos; no hay una guía metodológica para ello.

Pero la gestión de la variabilidad y reutilización en general, no es específica de la Ingeniería de Documentos. En Ingeniería del Software se aborda la variabilidad dentro de la Ingeniería de Líneas de Producto Software [4][5], que partiendo de una gestión explícita de la variabilidad, crea diferentes variantes de producto software. La aplicación de la Ingeniería de Líneas de Producto Software al dominio de la generación de documentos permite, desde nuestro punto de vista, proporcionar una guía metodológica al ingeniero de documentos, imprescindible cuando se enfrenta a dominios complejos, a la vez que permite integrar como parte de la propia línea, soluciones tecnológicas provenientes de la propia Ingeniería de Documentos. En este sentido, la herramienta DOPLER [6] ya combina ambas aproximaciones, aunque con diferencias con DPL, que comentaremos en las conclusiones.

3 Líneas de Producto de Documento

DPL propone la generación intensiva de familias de documentos. En [7] se presenta una primera aproximación a DPL, donde las líneas de documento producen editores personalizados para generar documentos, en base a su definición en términos de características. Las fases del proceso de desarrollo DPL son básicamente las mismas que definen las Líneas de Producto Software, como podemos ver en la Figura 1.

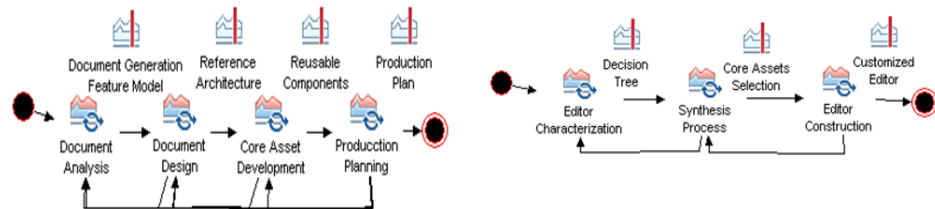


Fig. 1. El proceso de desarrollo de DPL en notación SPEM. Ingeniería del Dominio (parte izquierda) e Ingeniería de la Aplicación (parte derecha).

En la Ingeniería del Dominio, DPL establece una etapa de análisis de la familia de documentos, siendo la especificación de sus características, el principal artefacto a obtener, basado en el metamodelo propio de DPL, que se explicará en la siguiente subsección. Posteriormente, se define la arquitectura genérica del editor, identificando sus principales componentes software, entre los que destaca el repositorio que contendrá, entre otros, los fragmentos de documento. Finalmente, se especifica cómo dichos componentes son integrados de acuerdo a las relaciones definidas entre las características (el plan de producción). En la Ingeniería de la Aplicación, DPL sigue el marco de las Líneas de Producto Software, seleccionándose las características a incluir en el documento y generando el editor personalizado con los fragmentos de documento reutilizados. Finalmente, si se deben incluir datos variables dependientes del destinatario, en una segunda fase, se completará el documento con el editor.

3.1 Metamodelo de Características de una Familia de Documentos

La especialización del marco de Líneas de Producto Software, se basa en distinguir dos tipos de características basadas en los dos componentes de todo documento: contenido y presentación. El contenido incluye la estructura y los datos del documento. La presentación incluye, la tecnología para visualizarlo como documento impreso, pdf, video, etc. o formatos más ricos de representación como una aplicación multimedia o una aplicación web, más el diseño final del mismo. Estos componentes determinan la variabilidad de una familia de documentos, distinguiéndose variabilidad respecto al contenido y respecto a la presentación. En este último caso, sólo consideramos la variabilidad derivada de la tecnología utilizada, el diseño final del documento no se considera relevante en DPL, puesto que simplemente es una cuestión de ubicación de contenidos. En la Figura 2 se muestra el metamodelo.

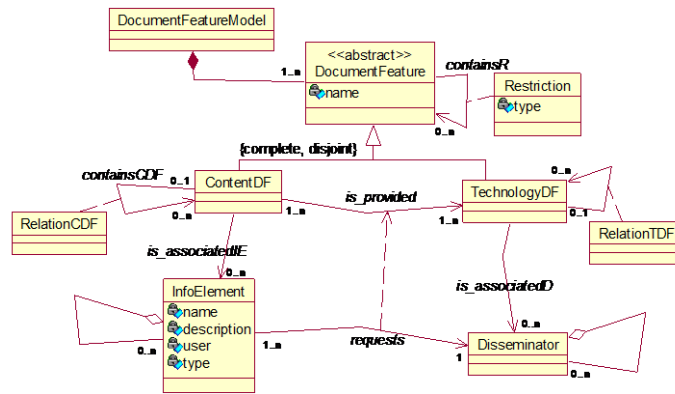


Fig. 2. Metamodelo de Características de una Familia de Documentos.

La principal aportación respecto a los modelos clásicos como FODA u otros posteriores, es la especialización en dos tipos de características: las relacionadas con el contenido del documento (*ContentDF*) y las relacionadas con la tecnología utilizada para representar dichos contenidos (*TechnologyDF*). Además, es importante la relación que existe entre ambas, puesto que cualquier característica de contenido para ser visualizada, necesita tener asociada una tecnología concreta. Por otra parte, el metamodelo propuesto se refina más al representar los fragmentos de contenido como instancias de una clase genérica denominada *InfoElement*; están asociados a las características de contenido y son almacenados como assets en el repositorio de DPL. De forma similar, cada característica tecnológica tiene asociadas un conjunto de soluciones tecnológicas para visualizar los contenidos, denominados genéricamente diseminadores (editores texto, visor de imágenes, de video, servicios web, etc.).

Los *InfoElements* deben ser representados, gestionados y recuperados como parte de la librería DPL de fragmentos de documento reutilizables. Usamos DITA como estándar para su implementación, estableciendo la equivalencia con sus tipos estructurales (tópicos y mapas), y definiendo los metadatos necesarios para su correcta recuperación y reutilización. DITA está concebido para generación de documentación técnica, como manuales, pero pensamos que es posible su utilización

para generar cualquier tipo de documento, gracias a la especialización de tipos y de dominios. Por otra parte, DITA OpenToolKit es la herramienta para la generación del documento final, lo que nos permite simular la funcionalidad proporcionada por los diseminadores asociados a las características tecnológicas. Debido a las limitaciones de espacio, no incluimos más detalles de la solución planteada con DITA.

4 Caso de Estudio: Generación de Planes de Emergencia

Los Planes de Emergencia son documentos complejos con información referente a la organización: riesgos, sistemas de alerta, procedimientos de actuación, mapas, etc. Su contenido varía en función del tipo de organización y su presentación se puede generar en varios formatos. La Figura 3 muestra el modelo de características. El marco de soporte a DPL se está desarrollando con Pure::Variants en la plataforma Eclipse. Algunos resultados se ilustran en la Figura 4(a), la arquitectura de referencia y los componentes del repositorio; o en la Figura 4(b), con un fragmento de plan de emergencia generado con DITA. Más detalles del caso de estudio se muestran en [8].

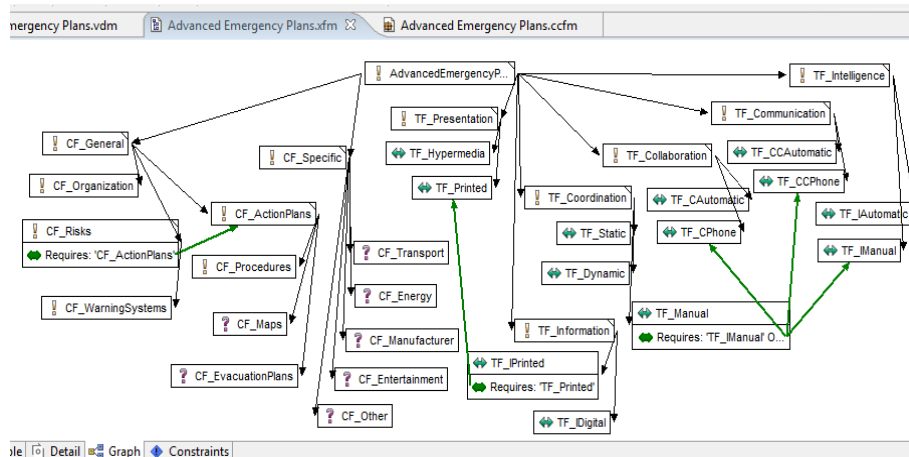


Fig. 3. Modelo de Características para los Planes de Emergencia.

5 Conclusiones y Trabajos Futuros

Se ha presentado la metodología Líneas de Producto de Documento (DPL) para la generación de familias de documentos. Se han definido puntos de variabilidad específicos del dominio (contenido y tecnología) que guían el proceso de producción del documento final, a partir del editor personalizado generado. Esta es una diferencia con la propuesta DOPLER. Son numerosas las líneas de actuación abiertas, entre las que destacan: la implementación de un marco de soporte a todo el proceso DPL; la utilización de DITA para la gestión de fragmentos de documentos, mucho más

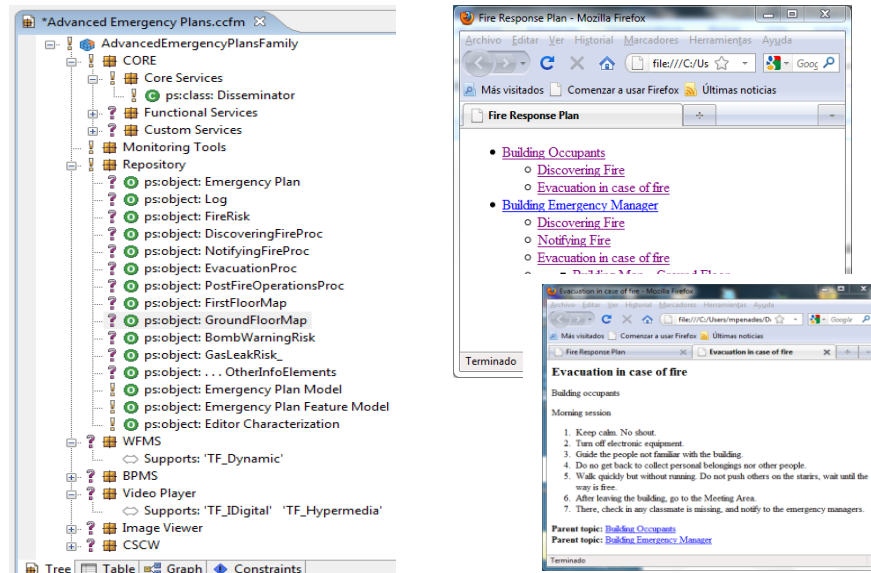


Fig. 4 (a). Arquitectura de referencia del editor de planes en DPL. **(b).** Ejemplo de Plan de Emergencia ante Incendio como documento HTML generado con DITA

flexible que DocBook, utilizado por DOPLER; o, la generación completa del editor. DPL se está validando en el dominio de generación de Planes de Emergencia.

Agradecimientos. M.C. Penadés y J.H. Canós ha sido financiados por el MEC-TIPEX (TIN2010-19859-C03-03). M. Borges ha sido financiado por CNPq-Brasil (No. 560223/2010-2 y 480461/2009-0). A. Vivacqua ha sido financiada por FAPERJ-Brasil. CAPES/MECD, Proyecto #169/ PHB2007-0064-PC ha financiado la cooperación entre ambos grupos.

Referencias

1. Lumley, J., Gimson R., Rees, O., A Framework for Structure, Layout & Function in Documents. Proc. 5th ACM Symposium on DocEng (2005)
2. Quint, V., Vatton, I., Editing with Style. Proc 7th ACM Symposium on DocEng (2007)
3. OASIS. Darwin Information Typing Architecture (DITA) Versión 1.2 (2010)
4. Clements, P., Northrop, L., Software Product Lines. Addison-Wesley (2002)
5. Pohl, K., B Böckle, G. and van der Linden, F., Software Product Line Engineering – Foundations, Principles, and Techniques. Springer, Berlin, Heidelberg, New York. (2005)
6. Rabiser, R., Heider, W., Elsner, C. et al. A Flexible Approach for Generating Product-Specific Documents in Product Lines. Proc. Software Product Line Conference (2010)
7. Penadés, M.C., Canós, J.H., Borges, M., Document Product Lines: Variability-driven Document Generation, Proc. 10th ACM Symposium on DocEng (2010)
8. Penadés M.C., Canós, J.H., Borges, M., Vivacqua, A., A Product Line Approach to the Development of Advanced Emergency Plans, Proc. International Conf. ISCRAM (2011)