Descubrimiento de Servicios Geográficos a partir de Casos de Uso Textuales

Viviana E. Saldaño Proyecto de Investigación Área Ingeniería de Software Unidad Académica Caleta Olivia - Universidad Nacional de la Patagonia Austral vivianas@uaco.unpa.edu.ar

Agustina Buccella, Alejandra Cechich Grupo de Investigación en Ingeniería de Software del Comahue (GIISCo) Departamento de Ciencias de la Computación Universidad Nacional del Comahue {abuccel, acechich}@uncoma.edu.ar

Resumen. El Desarrollo de Software Basado en Componentes (DSBC) se basa en la utilización de componentes software ya desarrollados OTS (Off-The-Shelf), que son combinados adecuadamente para satisfacer los requisitos de un sistema particular. Sin embargo, el uso masivo de este paradigma en la industria requiere de la identificación y selección de componentes de manera eficiente, aspecto que continúa siendo materia de investigación.

En este trabajo se da continuidad a trabajos previos, en los que hemos presentado una taxonomía de servicios geográficos que facilita la identificación de componentes, y es utilizada por el analista quien tiene a su cargo el desarrollo de una aplicación de un Sistema de Información Geográfica (GIS) y que utiliza para ello el enfoque de Desarrollo Basado en Componentes.

En esta instancia, se examinan distintas técnicas de extracción de conocimiento en casos de uso textuales y se propone un método para estandarizar el vocabulario resultante a fin de posibilitar la utilización de herramientas automáticas en la búsqueda de los servicios GIS correspondientes.

Palabras claves: DSBC, OTS, servicios GIS, taxonomías, casos de uso textuales.

1 Introducción

El Desarrollo de Software Basado en Componentes (DSBC) apela al uso de piezas pre-fabricadas, tal vez desarrolladas en momentos diferentes, por distintas personas y posiblemente con distintos objetivos de uso [22, 24]. En este contexto de desarrollo, cobran especial importancia los procesos de búsqueda y selección de componentes OTS [4]. Estos presentan serias limitaciones, tales como no documentación suficientemente expresiva de los componentes que garanticen una selección efectiva del mismo y además no contar con procesos de mediación que permitan agilizar la tarea de localización de componentes que brinden los servicios requeridos.

En este contexto, un cliente que requiere un servicio de un componente determinado, puede interrogar a un servicio de mediación para que éste le conteste con las referencias de aquellos componentes que proporcionan la clase de servicio requerido. Por otra parte, con el surgimiento del desarrollo de software basado en componentes, numerosas empresas fabricantes de GIS han comenzado a comercializar distintos tipos de componentes software orientados a las necesidades de los desarrolladores GIS. Para lograr un desarrollo más eficiente, los analistas se concentran en los atributos de reusabilidad e interoperabilidad. Sin embargo, se pierde mucho tiempo y esfuerzo en encontrar aquellos componentes que satisfagan la funcionalidad que se pretende implementar. Una de las necesidades clave para facilitar esta tarea, consiste en contar con información estándar de los componentes que permita agilizar la búsqueda de composiciones de software. De esta manera, la oferta de servicios debería poder estandarizarse de manera tal que las composiciones puedan formar parte de un repositorio de fácil acceso, así como también, la demanda de servicios debería poder expresarse en términos estándar para facilitar la búsqueda. Estos aspectos son los que determinan finalmente el éxito de un proceso de selección.

Este trabajo se presenta como una extensión al trabajo presentado en [20], en el cual se ha creado una taxonomía de servicios geográficos para facilitar la identificación de componentes. Además este trabajo se presenta como contraparte al modelo de oferta presentado en [8,9] en donde se define un servicio de publicación para facilitar la selección de los componentes solicitados. En este artículo, se propone un método para estandarizar el vocabulario utilizado por los analistas encargados de las aplicaciones GIS a fin de ayudar a la formulación de requerimientos para la búsqueda de componentes que provean los servicios GIS correspondientes.

Este artículo está organizado de la siguiente manera. A continuación en la Sección 2, se presentan los trabajos relacionados. Luego, en la Sección 3 se describe brevemente nuestra taxonomía de servicios geográficos y se define el método propuesto para encontrar los servicios requeridos partiendo de una especificación textual de casos de uso. En la Sección 4 se muestra una aplicación del método propuesto a un caso de uso real. Finalmente en la última sección se presentan las conclusiones y trabajo futuro.

2 Trabajos Relacionados

2.1 Casos de Uso

Los casos de uso han logrado una amplia aceptación como medio para describir las interacciones realizadas entre el sistema y su entorno. Actualmente el Lenguaje Unificado de Modelado (UML) [17] es un estándar altamente adoptado que define los conceptos centrales para el modelado de casos de uso. Al aplicar este estándar, los casos de uso se identifican y estructuran mediante los diagramas de casos de uso de UML y luego se especifican con descripciones textuales.

Existen distintas propuestas en cuanto a la notación utilizada para la especificación de casos de uso. La diferencia entre las mismas está relacionada con el nivel de formalismo empleado para realizar la especificación. Existen notaciones formales

[12, 21, 25], que proveen la posibilidad de inspeccionar y analizar las descripciones de casos de uso automáticamente, pero cuentan con la desventaja de la dificultad de interpretación para usuarios no técnicos. Existen también enfoques [2, 5, 9] que proponen la utilización de formas restringidas de lenguaje natural, mediante la utilización de plantillas para detallar el comportamiento de los casos de uso, de las cuales la más difundida es la propuesta de Cockburn [5]. Otra de las alternativas dentro de este enfoque es la utilización de lenguajes controlados, en la cual se restringe la gramática del lenguaje natural al uso de sólo ciertas palabras y estructuras de oración. Una de las propuestas de este enfoque, es la utilización del patrón SVDPI (Subject, Verb, Direct object, Preposition, Indirect object): "La estructura de la oración debe ser simple"..."Sujeto ... verbo ... objeto directo...preposición...objeto indirecto"[10].

En [11], se propone un metamodelo para descripciones textuales de casos de uso, que permite especificar el comportamiento de los mismos en forma de flujo. Este metamodelo define una representación textual del comportamiento de los casos de uso que preserva los beneficios del lenguaje natural, como la facilidad de lectura y entendimiento, para sus lectores contando además con el grado de formalidad requerido para realizar actividades de chequeo de consistencia entre ambas partes del modelo de casos de uso de UML: el modelo propiamente dicho y las descripciones textuales del comportamiento de los mismos.

Una alternativa interesante es la presentada por Bittner y Spence [3], en la cual se introduce cierto grado de formalidad para las descripciones de casos de uso y se permite a la vez la utilización del lenguaje natural sin restricciones. Esta propuesta consiste en describir los casos de uso en forma de flujos en lenguaje natural, los cuales definen el comportamiento a través de secuencias de eventos, contando con la ventaja de legibilidad para usuarios no técnicos.

En este trabajo utilizaremos un enfoque semi-formal, que permite aprovechar la ventaja de legibilidad del lenguaje natural y a su vez permite también que el texto pueda ser interpretado con herramientas automatizadas. Este enfoque es el de utilización de plantillas para la descripción de casos de uso, propuesto por Cockburn [5], en combinación con la utilización de lenguajes controlados, mediante la utilización del patrón SVDPI [10]. Del relevamiento realizado a organizaciones regionales, desarrolladoras y usuarias de aplicaciones geográficas, surge que no cuentan con documentación estandarizada que especifique los requerimientos de esas aplicaciones. Es por ello, que partiendo de especificaciones en lenguaje natural, se han elaborado los casos de uso textuales correspondientes, con el propósito de poder extraer el conocimiento contenido en los mismos, aprovechando las ventajas de este enfoque semi-formal.

2.2 Extracción de Conocimiento en Casos de Uso

Con respecto a la extracción de conocimiento en casos de uso, existen en la literatura una serie de trabajos. En [6, 14] se propone un método para procesamiento de requerimientos textuales, el cual se basa en la identificación de los atributos principales de las acciones descriptas en cada paso de un caso de uso, mediante la utilización de herramientas lingüísticas disponibles que permiten obtener un árbol

sintáctico. Habiendo identificado las acciones, permite conocer las operaciones aceptadas y solicitadas por una entidad, posibilitando la definición de interfaces de un componente, o servicios requeridos.

En [19] se proponen una serie de guías para escribir descripciones 'bien formadas' de casos de uso. Se realiza un interesante análisis estructural y semántico sobre los casos de uso lo cual da origen a los fundamentos del enfoque, un conjunto de patrones y estructuras lingüísticas que constituyen las estructuras fundamentales y superficiales de los casos de uso.

En [18] se presenta un modelo formal simple, Generic UC View, con el objetivo de permitir razonar sobre el comportamiento. También se introduce la sintaxis de procases, que al contar con guías de transformación sencillas posibilita la traducción de casos de uso textuales (basados en un template) a pro-cases [18]. Pro-cases es una técnica formal que permite especificar comportamiento, contando además con la ventaja de facilidad de lectura a todo tipo de usuario.

En este trabajo utilizaremos el método propuesto en [6, 14], dado que su aplicación es muy simple, no requiere aprendizaje de técnicas o herramientas auxiliares y las herramientas lingüísticas necesarias están disponibles en la web, contando en muchos casos con versiones demo en línea. A su vez, el método fue modificado para adaptarse a los requerimientos específicos de los servicios GIS. El mismo será explicado en la sección siguiente.

3 Identificación de Servicios GIS

En esta sección presentamos primero una breve descripción de la taxonomía definida en [20] la cual fue elaborada mediante un análisis de trabajos previos en la literatura [1, 23], los estándares del OGC [16] y un relevamiento efectuado a usuarios GIS pertenecientes a organizaciones de la zona acerca de sus expectativas sobre la funcionalidad de un sistema GIS.

Luego se describe la metodología propuesta para extraer conocimiento de los casos de usos textuales utilizados para la definición de los servicios requeridos por los usuarios GIS. El propósito final es instanciar nuestra taxonomía de manera de descubrir los servicios solicitados dentro de cada caso de uso para luego poder seleccionar los componentes GIS que deben ser requeridos para cumplir con los mismos.

3.1 Taxonomía de Servicios GIS

Para la generación de la taxonomía de servicios GIS, en [20] se tuvieron en cuenta los niveles o capas definidos por la ISO 19119[15] y los servicios ejemplo de cada una de ellas. A esta taxonomía se le anexó una capa de Requisitos no Técnicos, dado que en la selección de componentes OTS su consideración es tan importante como la correspondiente a la funcionalidad.

Para facilitar el proceso de mapeo entre los servicios encontrados a partir de los casos de uso y los servicios clasificados en la taxonomía de servicios GIS, se propone adicionar a dicha taxonomía, una columna con servicios ofrecidos, obtenidos

aplicando a la columna de descripción de ejemplos de servicios, el mismo método que se aplicó para la obtención de servicios a partir de casos de uso.

En la Tabla 1, se observa un fragmento de la taxonomía modificada, correspondiente a la capa Interacción Humana.

Tabla 1. Fragmento de Taxonomía de Servicios modificada

Capa	Servicio	Descripción Servicio	
		Verbo Principal	Objeto Representativo
Interacción Humana	Visor de catálogo	Localizar Explorar Gestionar	Metadato de dato geográfico Metadato de servicio geográfico
	Visor geográfico	Visualizar Mostrar Superponer Consultar Interactuar	Dato Capa Mapa
	Editor de servicios	Controlar Entender Componer Invocar Planificar	Servicio Cadena de servicio
	Editor de características geográficas	Interactuar Mostrar Consultar Agregar	Característica Orientación Perspectiva Profundidad Iluminación Transparencia Textura

3.2 Método Propuesto para la Extracción de Conocimiento en base a Casos de Uso

En [6, 14] se propone el método elegido para procesamiento de requerimientos textuales, el cual se basa en la identificación de los atributos principales de las acciones descriptas en cada paso de un caso de uso. Para esto se utilizan herramientas lingüísticas disponibles que permiten obtener un árbol sintáctico. Habiendo identificado las acciones, permite conocer las operaciones aceptadas y solicitadas por una entidad, posibilitando la definición de interfaces de un componente, o servicios requeridos.

El método presentado en [6, 14] parte de una especificación de casos de uso, construida mediante la utilización de las plantillas sugeridas por Cockburn y de las

_

¹ La capa Interacción Humana, en su versión completa cuenta con 10 Servicios definidos.

recomendaciones para escritura de casos de uso [5, 13], donde cada uno de los pasos que conforman el caso de uso debe cumplir con las siguientes premisas:

Premisa 1: Un paso de un caso de uso textual describe: (a) la comunicación entre un actor y el sistema, o (b) una acción interna

Premisa 2: La acción se describe mediante una oración simple, según el patrón SVDPI ("Sujeto...verbo...objeto directo...preposición...objeto indirecto")

Además el método consta de los siguientes pasos:

- 1) Determinación del POS (part-of-speech) de cada palabra, es decir el tipo de palabra y el rol que tiene dentro de la estructura de la frase.
- 2) Generación del árbol sintáctico de cada oración, es decir de cada paso que conforma el Escenario Principal del caso de uso
- 3) Obtención de los Tokens de Eventos. Estos tokens de eventos se construyen a partir del verbo principal de la oración y el objeto representativo (objeto directo).
- 4) A partir de los resultados obtenidos en el paso anterior, se convierten los casos de uso textuales en una especificación de comportamiento utilizando Pro-cases.

En este trabajo, hemos introducido algunas modificaciones al método explicado anteriormente para adaptarlo a las necesidades concretas del objetivo del mismo. En primer lugar, a diferencia de las herramientas lingüísticas utilizadas por el método original, en este trabajo se utiliza la suite de herramientas Freeling [7], que cuenta con soporte para idioma español. Otra variante introducida, consiste en cambiar el paso 4) del método original por otro proceso que permita encontrar un nombre de servicio estandarizado, definido en la Taxonomía de Servicios GIS, que sea funcionalmente equivalente al servicio obtenido mediante la generación del token de evento. Este proceso será realizado mediante la aplicación de una herramienta semi-automática que estamos desarrollando en la cual utilizaremos EuroWordNet2 como tesauro para el idioma español. Dicha herramienta es guiada por un usuario experto encargado de obtener y elegir relaciones de sinónimos adecuadas entre el token de evento y el servicio estándar de nuestro framework. La Figura 1 muestra gráficamente los pasos principales del método con las modificaciones previamente explicadas.

Como podemos observar, el método posee como entrada un plantilla de caso de uso textual y comienza con el análisis de la primera oración del Escenario Principal del caso de uso, que representa la primera acción del curso normal del mismo. Este análisis es realizado para cada una de las palabras de la oración, y consiste en la determinación del POS (part-of-speech), es decir tipo de palabra (verbo, sustantivo, etc) y el rol que cumple dentro de la oración. En el segundo paso del método, se genera el árbol sintáctico de la oración, en el cual las hojas son las palabras y los nodos son las frases.

A continuación, en el tercer paso del método se construye el token de evento, teniendo en cuenta el verbo principal y el objeto directo de la oración. En el paso 4, teniendo en cuenta la Taxonomía de servicios GIS modificada (Tabla 1), se realiza una comparación entre cada token de evento generado para cada acción del caso de uso y las descripciones de servicio de la Taxonomía. El desarrollador, con la utilización de la herramienta que utiliza EuroWordNet deberá seleccionar aquella descripción que semánticamente sea más aproximada al token de evento, y a partir de esta selección obtendrá el nombre estandarizado del servicio requerido y la capa a la que el mismo

² http://www.illc.uva.nl/EuroWordNet/

pertenece. Este resultado se almacena en un fichero XML, y constituirá el elemento base del proceso de demanda, mediante el cual se realiza la interrogación al servicio de mediación para que este proporcione las referencias de los componentes que brindan el servicio buscado.

Finalmente, todos estos pasos, 1-4, se repiten para cada una de las acciones que forman parte del Escenario Principal de la especificación del caso de uso.

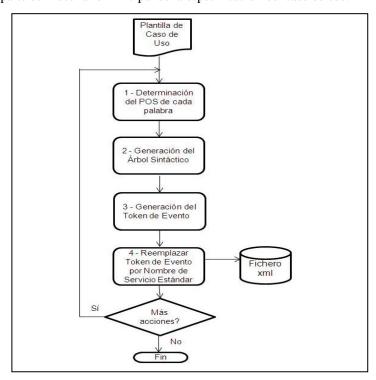


Figura 1. Método de extracción de servicios geográficos a partir de casos de uso.

4 Caso de Estudio

En esta sección presentamos un caso de uso textual perteneciente a una empresa de la zona de Comodoro Rivadavia. El mismo fue solicitado a través del relevamiento de servicios GIS efectuado a varias empresas y fue adaptado para enmarcarse en la plantilla de casos de uso definida en [5]. La empresa que proporcionó dicho requerimiento tenía los casos de usos definidos en forma textual sin ningún estándar. Junto con el analista de la empresa se enmarcaron algunos de ellos utilizando dicha plantilla. La Tabla 2 muestra uno de ellos, llamado "Modificar las coordenadas de ubicación de una línea eléctrica".

Tabla 2. Ejemplo Caso de Uso Textual

Nombre Caso de Uso	Modificar las coordenadas de ubicación de una línea elécrica		
Servicio	Actualizar la información de las líneas eléctricas		
Breve descripción	Asignar una nueva coordenada geográfica		
Actores	Administradores del sistema		
Alcance	Capa de tipo línea, representando a las líneas eléctricas		
Precondición	La capa de líneas eléctricas ha sido creada en el sistema.		
	1 El usuario selecciona una línea eléctrica		
Escenario Principal	2 El usuario modifica el atributo coordenada		
	Bl sistema muestra la línea eléctrica con ubicación actualizada		
Postcondición	El sistema tiene almacenada la ubicación actual de la línea eléctrica.		

De acuerdo al método aplicado, se analiza cada una de las acciones del Escenario Principal. En nuestro caso, al utilizar la suite de herramientas Freeling[7], los pasos 1 y 2 del método aplicado se realizan automáticamente en un solo proceso. En la figura 2 se muestra el árbol sintáctico obtenido a partir de la primera acción del Escenario Principal del caso de uso.

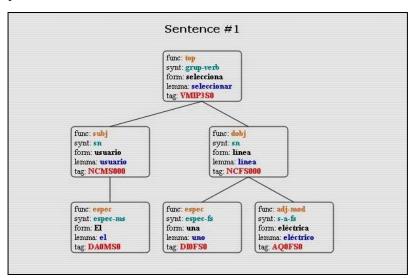


Fig. 2. Árbol Sintáctico de la oración: "El usuario selecciona una línea eléctrica"

Según propone el método en el paso 3, el token de evento, se conforma por el verbo principal y un objeto representativo, en sus formas base o lema.

El verbo principal es la palabra principal (raíz) de la frase verbal de nivel superior, en nuestro ejemplo: "selecciona". El objeto representativo se constituye a partir del

objeto directo de la oración, el cual se identifica como la primera frase sustantiva subordinada al verbo principal, en nuestro ejemplo: "línea eléctrica". Por lo tanto, el token de evento, para el primer paso del caso de uso, quedaría conformado como: "seleccionar línea eléctrica".

Finalmente como parte del paso 4, teniendo en cuenta nuestra taxonomía de servicios GIS, se realiza una comparación entre cada token de evento generado para cada paso del caso de uso y las descripciones de servicio de la Taxonomía. El desarrollador deberá seleccionar aquella descripción que semánticamente sea más aproximada al servicio requerido, y a través de esta selección obtendrá el nombre estandarizado del servicio requerido y la capa a la que el mismo pertenece.

En nuestro ejemplo, para encontrar el nombre estandarizado del servicio correspondiente al token de evento "seleccionar línea eléctrica" el usuario experto utilizará la herramienta bajo desarrollo para encontrar sinónimos entre las palabras del token de evento y las palabras dentro de la descripción del servicio en la taxonomía. En este caso, la herramienta junto con el usuario razonan que el verbo "seleccionar", en este contexto, es equivalente al verbo "mostrar" y que el objeto "línea" se corresponde con el objeto "característica". Así, la herramienta puede proponer la selección de un servicio estandarizado requerido, en este caso, el servicio seleccionado es "Editor de Características Geográficas" de la capa "Interacción Humana". Esta selección se almacena en un fichero XML.

5 Conclusión y Trabajos Futuros

En este trabajo se continúa con la tarea de facilitar el proceso de búsqueda de componentes geográficos. Para ello, y teniendo en cuenta que una de las herramientas más ampliamente utilizadas para el modelado de software es UML, se realizó un estudio sobre la información provista por la especificación de casos de uso, y las formas de utilizar esa información para encontrar los servicios que satisfagan la funcionalidad requerida en dichos casos de uso.

Se analizó y adaptó a la problemática particular de la búsqueda de servicios geográficos, un método de extracción de conocimiento en casos de uso. Para ello fue necesario también, adecuar la Taxonomía de Servicios Geográficos, definida anteriormente.

Como trabajo futuro, se trabajará sobre la automatización del método propuesto para su validación. Una vez realizado esto se combinará con la metodología definida en el servicio de publicación de servicios GIS [8,9] para realizar un mapeo entre oferta y demanda en forma semiautomática.

Referencias

- 1. Albrecht, J. Universal GIS Operations for Environmental Modeling, 1996.
- Armour, F., Miller, G. Advanced Use Case Modeling Volume One, Software Systems. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.
- 3. Bittner, K., Spence, I. Use Case Modeling. Addison-Wesley, 2000.

- Cechich A., Réquilé A., Aguirre J., Luzuriaga J. Trends on COTS Component Identification. 5th International Conference on COTS-Based Software Systems. Orlando, USA. IEEE Computer Science Press. 2006.
- Cockburn, A. Writing Effective Use Cases, Addison-Wesley Pub Co, ISBN: 0201702258, 1^{ra} ed., Enero 2000.
- Drazan, J., Mencl, V. Improved Processing of Textual Use Cases: Deriving Behavior Specifications. In Proceedings of SOFSEM 2007. LNCS, vol. 4362. pp. 856-886. Harrachov, Czech Republic, Enero 20-26, 2007.
- 7. Freeling Home Page, http://garraf.epsevg.upc.es/freeling/
- Gaetan, G., Cechich, A. Buccella, A. Un Esquema de Clasificación Facetado para Publicación de Catálogos de Componentes SIG. In Proceedings of CACIC 2008. Chilecito, La Rioja, Argentina, Octubre 2008. ISBN 987-24611-0-2
- Gaetan, G., Cechich, A. Buccella, A. Aplicación de Técnicas de Procesamiento de Lenguaje Natural y Web Semántica en la Publicación de Componentes para SIG. In Proceedings of ASSE 2009. Mar del Plata, Argentina, Agosto de 2009.
- 10. Graham, I.: Object-Oriented Methods: Princeiples and Practice, Addison-Wesley, 2000.
- 11. Hoffmann, V., Lichter, H., NyBen, A., Walter, A. Towards the Integration of UML-and textual Use Case Modeling. Journal of Object Tecnology, Vol. 8, No. 3, 2009.
- Kholkar, D., Krishna, G., Shrotri, U., and Venkatesh, R. Visual Specification and Analysis of Use Cases. In SoftVis '05: Proceedings of the 2005 ACM symposium on Software visualization, pp 77-85, New York, NY, USA, ACM, 2005.
- 13. Kulak, D., Guiney, E. Use Cases:Requirements in Context. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, ISBN: 0321154983, 2da ed., Agosto 2003.
- 14. Mencl, V. Deriving Behavior Specifications from Textual Use Cases In: Proceedings of Workshop on Intelligent Technologies for Software Engineering (WITSE04, part of ASE 2004), Linz, Austria, Oesterreichische Computer Gesellschaft, Sept. 2004.
- 15. Percivall, G. ISO 19119 and OGC Service Architecture. FIG XXII International Congress, Washington, D.C. USA, 2002.
- 16. OGC. Topic 12: OpenGIS Service Architecture. Open GIS Consortium, 2002
- OMG. UML Superstructure Specification, v2.1.2. OMG Formal Document 2007-11-02, Noviembre 2007.
- Plasil, F., Mencl, V. Getting "Whole Picture" Behavior in a Use Case Model. Transactions of SDPS: Journal of Integrated Design and Process Science 7(4), pag.63-79, Dia 2003
- Rolland, C, Achour, C. Guiding the construction of textual use case specifications. In Data & Knowledge Engineering, vol. 25 no. 1-2, pág. 125-160, Marzo 1998.
- Saldaño, V., Buccella, A., Cechich, A. Una Taxonomía de Servicios Geográficos para facilitar la identificación de componentes. In Proceedings of CACIC 2008. Chilecito, La Rioja, Argentina, Octubre 2008. ISBN 987-24611-0-2
- 21. Spivey, J. The Z Notation: A Reference Manual. Prentice Hall, 1992.
- 22. Szyperski, C. Component Software-Beyond Object-Oriented Programming. Addison-Wesley, 1998.
- 23. Timpf, S. Geographic Task Models for Geographic Information Processing. Meeting on Fundamental Questions in Geographic Information Science, Manchester, UK. 2001
- Wallnau, K., Hissam, S., and Seacord, R. Building Systems from Commercial Components. Addison-Wesley, 2002.
- 25. Whittle, J and Jayaraman, P. Generating Hierarchical State Machines from Use Case Charts. In RE '06: Proceedings of the 14th IEEE International Requirements Engineering Conference (RE'06), pp 16-25. Washington, DC, USA, IEEE Computer Society, 2006.