

Selección de Componentes para SIG a partir de Información en Servicios y Catálogos Web Normalizados

Gabriela Gaetan, Adriana Martin, Sonia Molina, Viviana Saldaño
Unidad Académica Caleta Olivia
Universidad Nacional de la Patagonia Austral
Caleta Olivia, Santa Cruz, Argentina

Agustina Buccella, Alejandra Cechich
GHISCo, Departamento de Ciencias de la Computación
Universidad Nacional del Comahue
Neuquén, Argentina

CONTEXTO

Este artículo describe las actividades que se están desarrollando en el proyecto “Mejora del Proceso de Selección de Componentes para Sistemas de Información Geográficos” (29/B107), el cual es continuación del proyecto “Identificación de Servicios para Sistemas de Información Geográficos (SIG)” (29/B090).

RESUMEN

Uno de los problemas en la recuperación de componentes de software existentes es la diferencia entre la información que el cliente especifica sobre el componente que solicita y la información contenida en las descripciones de componentes. Además, en el caso particular de un componente SIG (Sistemas de Información Geográficos), dado que las aplicaciones son utilizadas en dominios muy diferentes, cada una de ellas posee un punto de vista distinto sobre la información geográfica, provocando que en el proceso de desarrollo de aplicaciones SIG, sea muy complejo encontrar la información adecuada sobre los componentes requeridos. Por otra parte, la descripción de componentes encontrada en los catálogos web contiene tanto información comercial, como información técnica y funcional. Sin embargo, cuando el cliente especifica la consulta lo hace enfocado principalmente en la funcionalidad del componente, y con mayor nivel de detalle. En este trabajo, nos centramos en mejorar los procesos de selección y publicación de componentes, proponiendo metodologías que mejoren la forma de realizar las tareas incluidas en ellos.

Palabras clave: Desarrollo OTS; Sistemas de Información Geográficos; Procesamiento de Lenguaje Natural; Desarrollo Basado en Componentes.

1. INTRODUCCION

El Desarrollo de Software Basado en Componentes (DSBC) procura reducir el tiempo de desarrollo, los costos y el esfuerzo, y también mejorar la calidad del producto final, por medio del reuso de componentes que pueden ser recuperados y ensamblados en aplicaciones específicas. Para explotar la reusabilidad de los

componentes disponibles debe ser posible encontrar los componentes más apropiados de acuerdo a los requerimientos de los usuarios. Por lo tanto, durante el proceso de selección, es deseable contar con un repositorio que utilice metadatos para describir las características de los componentes ofrecidos a la comunidad de usuarios.

En la literatura se encuentran distintos aportes relacionados con la clasificación, búsqueda y recuperación de componentes [2]. Algunos trabajos [1, 3, 5, 8, 9, 11], proponen resolver los problemas relacionados con la categorización (o indexación), concentrando sus propuestas principalmente en las estructuras de clasificación de la información sobre componentes. Otros trabajos, además de realizar propuestas de clasificación para componentes, avanzan en la solución de los problemas de búsqueda y almacenamiento de información sobre los componentes. Por ejemplo, el sitio web Componex, permite realizar la búsqueda en su repositorio al organizar y clasificar los componentes de software, usando un esquema XML bien definido para especificar distintos parámetros (tecnología, plataforma, dominio, etc.). ComponentXchange [12] es una infraestructura enfocada tanto en la especificación como en la búsqueda y localización de los componentes que mejor satisfagan los requerimientos de los integradores. Los documentos de especificación de los componentes son enviados por los vendedores y se los almacena en un repositorio de descripción de componentes.

Otras propuestas, como en [10], proponen un enfoque basado en ontologías y modelos de dominio, orientado a incrementar la efectividad de la búsqueda y brindar información que relacione los componentes recuperados. En [4] se describe un sistema para clasificación y recuperación que extrae automáticamente, a partir de las descripciones del software, el conocimiento necesario para catalogar los componentes en la base de conocimiento usando técnicas de procesamiento de lenguaje natural. Todas las soluciones enumeradas se enfocan sólo en algún problema particular relacionado con la recuperación de información de componentes. Algunas soluciones están dirigidas a los aspectos relacionados con la clasificación, mientras que otras están enfocadas en

aspectos de documentación y almacenamiento de la información relacionada con los componentes OTS (Off-The-Shelf).

Por otra parte, con el surgimiento del desarrollo de software basado en componentes, numerosas empresas fabricantes de sistemas de información geográfica (SIG) han comenzado a comercializar distintos tipos de componentes software orientados a las necesidades de los desarrolladores SIG. Sin embargo, dado que las aplicaciones SIG son utilizadas en dominios muy diferentes, cada aplicación posee un punto de vista distinto sobre la información geográfica. Cada desarrollador ha definido modelos conceptuales, modelos de datos geográficos, formatos de almacenamiento, operaciones de análisis o procedimientos de representación especialmente adaptados a los requerimientos de la aplicación. Por lo tanto, en el proceso de desarrollo de aplicaciones SIG, encontrar información adecuada sobre componentes geográficos es una tarea crucial; los catálogos disponibles en la web ofrecen descripciones de los componentes, pero los mecanismos de selección, generalmente basados en búsqueda de palabra-clave, son poco efectivos.

En consecuencia, el DSBC para aplicaciones de dominio SIG, es un proceso complejo, en el que intervienen por un lado el desarrollador del componente y por otro, el usuario del componente, definiéndose así dos modelos, de demanda y de oferta, definiendo requerimientos especiales y soluciones de acuerdo a las restricciones y problemas que involucra cada uno de ellos.

El modelo de demanda hace referencia a los requerimientos o servicios solicitados por los clientes de SIG (empresas que desarrollan SIG), los cuales, en un desarrollo basado en componentes, buscan en la Web, componentes que brinden servicios que se adecuen a las necesidades de sus aplicaciones.

El modelo de oferta se refiere a los requerimientos o servicios que brindan los desarrolladores de componentes para SIG, para lo cual publican en la Web los componentes desarrollados junto a la información técnica y funcional asociada a los mismos.

Por lo tanto, para que ambas partes logren entenderse, es fundamental la forma en que los componentes son publicados, llamado proceso de publicación, y la forma en que los mismos son seleccionados, llamado proceso de selección (Figura 1). Además, para integrar la funcionalidad de ambos procesos surge la necesidad de utilizar un modelo de mediación que facilite tanto la recuperación de la información como la selección automatizada de componentes candidatos [6].

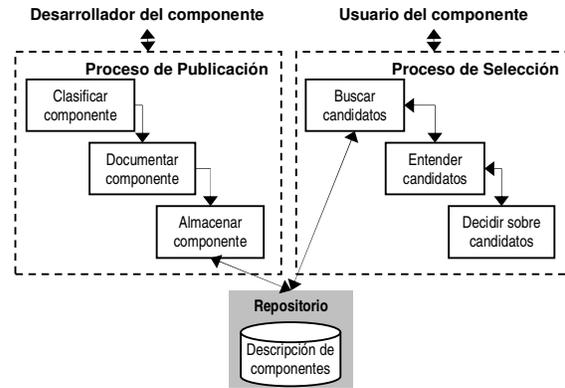


Figura 1. Proceso de selección de componentes

2. LINEAS DE INVESTIGACION y DESARROLLO

Hasta el momento, se ha trabajado sobre ambos modelos, de oferta y de demanda, proponiendo metodologías que mejoren la forma de realizar las tareas incluidas en ellos.

Para el modelo de demanda, con el propósito de mejorar el proceso de selección y búsqueda de componentes SIG, en [19] definimos una taxonomía de servicios basándonos en el estándar para servicios SIG de la ISO 19119 [7]. Además, teniendo en cuenta que una de las herramientas más ampliamente utilizadas para el modelado de software es UML, en [20] [21] se realizó un estudio sobre la información provista por la especificación de casos de uso y se propuso un método para extraer conocimiento en base a Casos de Uso textuales, que facilita la búsqueda de servicios que satisfacen la funcionalidad requerida en dichos casos de uso. Este método utiliza herramientas, las cuales se valen de técnicas de lenguaje natural, para encontrar la clasificación más adecuada del servicio requerido, según la taxonomía definida [19].

Para el modelo de oferta, en [14,16] definimos otra taxonomía conteniendo la información necesaria para describir un componente, y así ayudar al proceso de publicación. El objetivo principal de este trabajo fue analizar qué información está realmente presente en la documentación de los catálogos de componentes para saber con qué datos contar a la hora de implementar el modelo de mediación. En [17], describimos el proceso de Extracción de información enriquecido con tecnologías asociadas al Procesamiento de Lenguaje Natural y a la Web Semántica. La idea principal consiste en poblar automáticamente un esquema de clasificación normalizado (de aspectos funcionales y no técnicos) de componentes SIG con la información disponible en portales Web especializados.

Asimismo, para realizar el mapeo de la información provista por ambos modelos, oferta y demanda, se está trabajando sobre el desarrollo de una herramienta que facilite dicho mapeo [18]. Para ello es necesario resolver el problema de desambiguación del sentido de las palabras inherente al uso del lenguaje natural.

Por otra parte, los catálogos en la Web han evolucionado rápidamente hacia aplicaciones de negocios complejas y servicios online. Al igual que cualquier otro sistema interactivo de información, estas aplicaciones, no solo deben cumplir con los objetivos funcionales para las que han sido creadas, tarea que no es nada fácil, sino que además deben cumplir con todos los principios y características deseables, que permiten calificarlas como “accesibles”.

La norma ISO/TC 16027 [13], define Accesibilidad como la facilidad de uso en forma eficiente, eficaz y satisfactoria de un producto, servicio, entorno o instrumento por personas que poseen capacidades diferentes. Por lo tanto, Accesibilidad electrónica hace referencia a que los productos y servicios electrónicos puedan ser utilizados por los usuarios con efectividad, eficiencia y satisfacción en un contexto de uso determinado.

En el caso de una aplicación Web, que sea “accesible”, significa que debe presentar la información a las personas, de manera tal que independientemente de la tecnología que utilicen (ordenador, PDA, teléfono, etc.) y de las capacidades diferentes que posean (físicas, psíquicas, sensoriales u otras) estén en igualdad de condiciones en lo que al acceso a la información se refiere. Muchos países ya han legislado para velar por la Accesibilidad de los sitios relacionados con los organismos del estado, es decir aquellos referidos a la administración y provisión de servicios públicos a sus ciudadanos. Así mismo han surgido asociaciones y organizaciones en todo el mundo, que persiguen el objetivo de encaminar la Web a su máximo potencial, lo cual implica una Web para todos: un Acceso Universal. Por ejemplo, el World Wide Web Consortium (W3C) [14], fue creado en Octubre de 1994 con la visión “Universal Web Access: The Web Anywhere, for Everyone, at Anytime, on Everything”. La W3C ha desarrollado guías de Accesibilidad denominadas Web Content Accessibility Guidelines 1.0 (WCAG 1.0) [15] que son consideradas en la Unión Europea como normas de facto, y citadas como referencia obligada en la mayoría de las legislaciones sobre Tecnologías de la Información y Comunicación de todo el mundo. Estas recomendaciones explican cómo hacer “accesible” el contenido de la Web a personas con discapacidad y están pensadas tanto para los desarrolladores de contenido (autores de páginas y diseñadores de sitios) como para los desarrolladores de herramientas de autor. Basados en estas recomendaciones un gran número de enfoques han surgido en los últimos años y están disponibles para asistir a los desarrolladores Web en la generación de contenido “accesible”. En [16] se presentan y comparan quince de estos enfoques encontrados en la literatura reciente en un framework de evaluación denominado WAAM “Web Accessibility Assessment Model”. Considerando que acceder información de catálogos Web es un proceso inherente a la selección de componentes y que la importancia de definir sitios web accesibles está estrechamente ligada a

una selección exitosa, hemos decidido extender nuestra línea de investigación para abarcar este aspecto.

En síntesis, la oferta de servicios debería poder estandarizarse de manera tal que las composiciones puedan formar parte de un repositorio de fácil acceso. De manera similar, la demanda de servicios debería poder expresarse en términos estándar para facilitar la búsqueda. Estos aspectos son los que determinan el éxito finalmente de un proceso de selección y son el principal foco de investigación de este proyecto en el dominio de los sistemas de información geográficos.

3. RESULTADOS OBTENIDOS/ ESPERADOS

Actualmente, nuestro trabajo se centra en la instanciación del modelo de mediación, para lo cual es necesario definir un proceso mediador y la herramienta asociada que nos permita integrar los procesos de Publicación y Selección de componentes SIG. Para lograr este objetivo, se están analizando técnicas de Lenguaje Natural [4] y Web Semántica.

Por otra parte, será necesario realizar un modelo de evaluación (aspectos de calidad de la información recolectada, calidad de la solución encontrada) y recomendaciones de mejora de sitios Web de publicación de catálogos de componentes.

4. FORMACION DE RECURSOS HUMANOS

Este proyecto está integrado por 6 (seis) investigadores, entre los que se cuentan docentes de la Universidad Nacional del Comahue (UNC) y de la Universidad Nacional de la Patagonia Austral (UNPA).

El mayor impacto del proyecto se centra en la formación de recursos humanos. Actualmente 2 (dos) investigadores están realizando sus tesis de Maestría en la Universidad Nacional de La Plata (UNLP), bajo la supervisión de investigadores de UNC y UNLP; 1 (un) docente está realizando su tesis de Doctorado en la UNLP y 1 (un) docente está iniciando la Maestría en la Universidad Nacional de la Patagonia Austral.

5. BIBLIOGRAFIA

1. Ackermann, J., Brinkop, F., Conrad, S., Fettke, P., Frick, A., Glistau, E., Jaekel, H., Kotlar, O., Loos, P., Mrech, H., Ortner, E., Overhage, S., Raape, U., Sahm, S., Schmietendorf, A., Teschke, T., Turowski, K.: Standardized Specification of Business Components. German Society of Informatics (2002)
2. Cechich A., Réquile 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.
3. Dong, J., Alencar, P. S. C., Cowan, D. D.: A Component Specification Template for COTS-based Software Development. First Workshop on Ensuring Successful COTS Development (1999)
4. Girardi, M. R.; Ibrahim, B.: A software reuse system based on natural language specifications. 5th

International Conference on Computing and Information. Ontario (1993)

5. Iribarne, L., Troya, J. M., Vallecillo, A.: Trading for COTS Components in Open Environments. 27th Euromicro Conference. IEEE Computer Society Press. Warsaw (2001)
6. Iribarne, L. Un Modelo de Mediación para el Desarrollo de Software Basado en Componentes COTS. Phd thesis, Universidad de Almería. (2003)
7. ISO/IEC. Geographic Information Services. Draft International standard 19119 (2002)
8. Kallio, P., Niemelä, E.: Documented Quality of COTS and OCM Components. 4th ICSE Workshop on Component-Based Software Engineering (2001)
9. Prieto-Díaz, R., Freeman, P.: Classifying Software for Reusability. IEEE Software. 4(1):6-16 (1987)
10. Sugumaran, V.; Storey, V.C.: A semantic-based approach to component retrieval. SIGMIS Database. 34(3): p. 8-24 (2003)
11. Torchiano, M., Jaccheri, L., Sørensen, C., Wang, I.: COTS Products Characterization. 14th international conference on Software engineering and knowledge engineering. Ischia (2002)
12. Varadarajan, S.; Kumar, A.; Gupta, D.; Jalote, P.: ComponentXchange: An E-Exchange For Software Components. IADIS Conf. WWW/Internet. Portugal (2001)
13. ISO/TS 16071 International Organization for Standardization/Technical Specification. Ergonomics of Human-System Interaction - Guidance on Accessibility for Human-Computer Interfaces, 2002.
14. W3C Web Accessibility Initiative (WAI): Introduction to Web Accessibility. <http://www.w3.org/WAI/intro/accessibility.php>
15. WCAG 1.0 Web Content Accessibility Guidelines 1.0. World Wide Web Consortium (W3C) Recommendations, 1999. Disponible en <http://www.w3.org/TR/WAI-WEBCONTENT/>
16. Martín, A. Cechich, A and Rossi, G: Comparing Approaches to Web Accessibility Assessment. Handbook of Research on Web Information Systems Quality, Idea Group Inc., C. Calero, M^a Á. Moraga, M. Piattini (Eds.), ISSN: 0950-5849, 2008

CONTRIBUCIONES 2008-2011

1. Gaetán, G., Buccella, A., Cechich A., CLASIFICACIÓN DE COMPONENTES OTS (OFF-THE-SELF) PARA SISTEMAS DE INFORMACIÓN GEOGRÁFICOS, Actas del X Workshop de Investigadores en Ciencias de la Computación, ISBN 978-950-863-101-5, 2008.
2. Gaetán, G., Buccella, A., Cechich A., UN ESQUEMA DE CLASIFICACIÓN FACETADO PARA PUBLICACIÓN DE CATÁLOGOS DE COMPONENTES SIG, Actas del XIV Congreso

Argentino en Ciencias de la Computación, ISBN 987-24611-0-2,2008.

3. Gaetán, G., Buccella, A., Cechich A., Saldaño V., CLASIFICACION DE COMPONENTES SIG, Actas del XI Workshop de Investigadores en Ciencias de la Computación, ISBN 978-950-605-570-7, 2009.
4. Gaetán G. Cechich A., Buccella A., APLICACIÓN DE TÉCNICAS DE PROCESAMIENTO DE LENGUAJE NATURAL Y WEB SEMÁNTICA EN LA PUBLICACIÓN DE COMPONENTES PARA SIG, Actas del ASSE (Simposio Argentino de Ingeniería de Software), JAIHO, ISSN 1850-2792 (110—121), 2009.
5. Gaetán G., Cechich A., Buccella A., EXTRACCIÓN DE INFORMACIÓN A PARTIR DE CATÁLOGOS WEB DE COMPONENTES PARA SIG, Actas del XV Congreso Argentino en Ciencias de la Computación, ISBN 978-897-24068-4-1, (891—900), 2009.
6. Gaetan G., Saldaño, V., Buccella, A., Cechich, A., A DOMAIN-ORIENTED APPROACH FOR GIS COMPONENT SELECTION, The Fifth International Conference on Software Engineering Advances, ICSEA 2010, August 22-27, 2010 - Nice, France
7. Saldaño, V., Buccella, A., Cechich, A., UNA TAXONOMÍA DE SERVICIOS GEOGRÁFICOS PARA FACILITAR LA IDENTIFICACIÓN DE COMPONENTES, Actas del XIV Congreso Argentino en Ciencias de la Computación, ISBN 987-24611-0-2, 2008
8. Saldaño V., Buccella A., Cechich A., DESCUBRIMIENTO DE SERVICIOS GEOGRÁFICOS A PARTIR DE CASOS DE USO TEXTUALES, Actas del XV Congreso Argentino en Ciencias de la Computación, ISBN 978-897-24068-4-1, (921—930), 2009.
9. Saldaño V., Buccella A., Cechich A., DISCOVERING GEOGRAPHIC SERVICES FROM TEXTUAL USE CASES, Journal of Computer Science & Technology, Vol. 10 - No. 2 - June 2010 - ISSN 1666-6038
10. Gaetán, G., Buccella, A., Cechich A., Martín, A., Molina S., Saldaño V., RECUPERACIÓN DE INFORMACIÓN DE COMPONENTES SIG ENRIQUECIDO CON TECNOLOGÍAS SEMÁNTICAS. Encuentro de Investigadores de Ciencias Básicas, Aplicadas y Experimentales de Patagonia Austral EIBAE. 2010.
11. Martín, A., Cechich A., SUPPORTING AN ASPECT-ORIENTED APPROACH TO WEB ACCESSIBILITY DESIGN. The Fifth International Conference on Software Engineering Advances, ICSEA 2010, August 22-27, 2010 - Nice, France
12. Gaetán, G., Buccella, A., Cechich A., Martín, A., Molina S., Saldaño V., PUBLICACIÓN Y

SELECCIÓN DE COMPONENTES PARA SIG.
Actas del XII Workshop de Investigadores en
Ciencias de la Computación, 2010.

13. Martin, A., Cechich A., ACCESSIBILITY AT
EARLY STAGES: INSIGHTS FROM THE
DESIGNER PERSPECTIVE. 8th International
Cross-Disciplinary Conference on Web
Accessibility. Hyderabad, India. 2011.