

Definiendo un Marco Riguroso para Diseñar un Lenguaje de Modelado de Software Confiable y Consistente para la Web

Sonia Flores, Salvador Lucas, Roberto Uzal

*Departamento de Informática- Universidad Nacional de San Luis
Ejército de los Andes 950- San Luis Argentina
DSIC - Universidad Politécnica de Valencia
Camino de Vera s/n - Valencia, España
sflores@unsl.edu.ar, slucas@dsic.upv.es, ruzal@uolsinectis.com.ar*

8 de abril de 2010

Resumen

En esta Línea de Investigación, nos proponemos encontrar los mecanismos para conformar un marco riguroso dentro del cual abordar importantes fallas de calidad, de escalabilidad, de rendimiento (performance) y de seguridad (entendida como confidencialidad) que presentan los sitios Web y sus aplicaciones, con el fin de encontrar una formalización de las especificaciones de las aplicaciones Web, bajo la premisa de que un proceso de desarrollo confiable y consistente debe estar sustentado por modelos conceptuales que tengan la expresividad necesaria, para poder especificar las singularidades que encontraremos en el espacio de problemas de las aplicaciones Web. La Ingeniería del Software, aplicada en el contexto de las aplicaciones Web, requiere de la disponibilidad de lenguajes de alto nivel que permitan especificar con precisión las propiedades de interés, como así también resulta necesario disponer de algoritmos que permitan verificar dichas propiedades, una vez expresadas en el lenguaje adecuado.

La simplicidad de los lenguajes de especificación basados en reglas y la cercanía de éstos con determinadas clases de lógicas y autómatas, sugieren su empleo en las tareas de modelado de sitios Web.

1. Contexto

Esta propuesta de trabajo se desarrolla como tesis doctoral en Ciencias de la Computación y se espera que de la misma se definan otros trabajos de investigación que sirvan como futuras tesis de Maestría y tesis de Licenciatura.

2. Introducción

A pesar de que actualmente millones de usuarios hacen uso de las aplicaciones Web, del correo electrónico y de servicios de audio y de video, las prestaciones asociadas presentan importantes debilidades. Entre ellas se destacan importantes fallas de calidad, de escalabilidad, de rendimiento (performance) y de seguridad (entendida como confidencialidad). La formalización, con un enfoque riguroso, de las especi-

ficaciones de las aplicaciones Web, presenta importantes oportunidades de investigación y de desarrollo.

A efectos de acotar las posibles interpretaciones, entendemos como aplicación web a la que está implantada en un sitio Web donde, la navegación a través del mismo y la entrada de datos por parte de los usuarios, afectan el estado de la lógica del negocio. En esencia, una aplicación Web usa a un sitio Web como front-end. Acordamos que, si no existe lógica del negocio a nivel servidor, el sistema no puede ser llamado estrictamente aplicación Web. En general encontramos una capa intermedia, integrada por un conjunto de componentes, que se ejecutan no necesariamente en el servidor Web, sino en otros servidores de aplicaciones. Esta capa suele encapsular la lógica del negocio y, al ser componentes compilados, puede contener objetos, incluyendo sus métodos y atributos (*llamados business objects*).

Como se anticipó, importantes y trascendentes aportes conceptuales e instrumentales pueden concretarse en el ámbito de las aplicaciones Web. Por ejemplo, un proceso de desarrollo confiable y consistente debe de estar sustentado por modelos conceptuales que tengan la expresividad necesaria para poder especificar las singularidades que encontraremos en el espacio de problemas de las aplicaciones Web.

La Ingeniería del Software aplicada en el contexto de las aplicaciones Web requiere de la disponibilidad de lenguajes de alto nivel que permitan especificar con precisión las propiedades de interés. También resulta necesario disponer de algoritmos que permitan verificar dichas propiedades, una vez expresadas en el lenguaje adecuado.

La dualidad entre especificación lógica y determinados tipos de mecanismos formales de decisión (en particular, diversas clases de autómatas) proporciona un marco muy adecuado para abordar los problemas anteriores. La simplicidad de los lenguajes de especificación basados en reglas y la cercanía de éstos con determinadas clases de lógicas y

autómatas, profusamente investigadas en las últimas décadas, sugieren su empleo en las tareas de modelado de sitios Web. De hecho, los recientes intentos de aplicar dichas técnicas confirman este punto de vista.

3. Estado del Arte

La creciente complejidad de los sitios Web ha convertido el problema de la producción de sitios web con un buen desempeño en un verdadero reto. De hecho, es bastante más sencillo encontrar información inconsistente en la Web que un sitio debidamente mantenido en Internet. Por esta situación el diseño de sitios Web, la construcción y el mantenimiento son fases que deberían ser tratadas desde una perspectiva de ingeniería para poder proporcionar información consistente y fiable. A pesar de que la gestión de sitios Web ha recibido una especial atención a lo largo de los últimos años pocos trabajos han profundizado en el análisis y diseño de sitios Web. En [20], se desarrolla un algoritmo declarativo de verificación que comprueba una clase particular de restricciones de integridad que se refieren a la estructura del sitio Web, pero no a los contenidos de una instancia dada del sitio Web. En [17], se propone una metodología para verificar algunas restricciones semánticas referentes al contenido del sitio Web, que consiste en usar reglas de inferencia y axiomas de la semántica natural. En [18] se formaliza una aproximación simbólica de comprobación de modelos, la cual construye un modelo de estados finitos del sistema en el lenguaje de entrada del comprobador de modelos, y después comprueba las propiedades consideradas expresadas en lógica CTL. Para un examen comprensivo sobre el problema general de la comprobación de restricciones entre múltiples documentos, nos referimos a [21, 19, 16]. Creemos que una aproximación basada en maquinaria de tipo reescritura para la especificación y verificación de sitios Web puede ser empleada de forma muy eficiente, ya que puede aprovechar las capacidades del razonamiento típicas del mundo de la programación declarativa. Nuestra idea

es que las técnicas de reescritura de términos pueden dar soporte de una forma no solo intuitiva sino también natural a la especificación de alto nivel de sitios Web, y también a técnicas eficientes de verificación de sitios Web. Por lo que sabemos, los trabajos [2, 3] y [22] proporcionan una primera muestra de la utilidad de dichas técnicas que el trabajo desarrollado durante la realización de la presente línea de investigación extenderá y profundizará. Sólo tenemos conocimiento de dos aproximaciones basadas en reescritura para el procesamiento de páginas Web, pero se centran en tareas de transformación en lugar de en tareas de verificación: en [23] se facilita una implementación de la reescritura de términos para (un fragmento de) XSLT, el lenguaje basado en reglas diseñado por W3C para la transformación de documentos XML, mientras que en [15] se utilizan reglas de reescritura para realizar transformaciones HTML con el fin de mejorar las aplicaciones Web depurando su sintaxis, reorganizando los marcos o actualizando el portal a nuevos estándares. Los datos y los programas son los ingredientes básicos del software. Con el fin de resolver un problema, necesitamos programas que se comporten de forma correcta respecto a las intenciones del programador y que operen sobre conjuntos correctos de datos. Este hecho ha llevado a muchos investigadores a enfrentarse al problema del diseño, la verificación y corrección de datos y programas. En esta línea de investigación, intentaremos aportar nuestra pequeña contribución a este campo, centrándonos en el diseño.

4. Líneas de Investigación y Desarrollo

La presente Línea de Investigación está orientada a profundizar en esta perspectiva mediante:

- La definición de distintos lenguajes de modelado basados en reglas que permitan distintos tipos de representación de los sitios Web así como transformaciones correctas y completas de las distintas

representaciones que permitan pasar de una a otra según convenga. Dichas representaciones deberán cubrir tanto los aspectos estáticos (contenido, estructura de las páginas Web) como dinámicos (transiciones, evolución de los sitios Web, seguridad, etc.).

- La identificación de fragmentos decidibles de lenguajes lógicos que permitan especificar propiedades de los sitios Web, así como la implementación eficiente de los algoritmos de decisión que permitan verificar dichas propiedades, y que correspondan con alguno de los niveles de especificación indicados en el primer ítem.
- El desarrollo de análisis automáticos y transformaciones que permitan optimizar, corregir e informar sobre el estado de un sitio Web en función de determinadas especificaciones de calidad para los mismos.
- El desarrollo de mecanismos de certificación de sitios Web.

5. Resultados Esperados

En nuestra opinión, una aproximación formal, sistemática, puede proporcionar muchos beneficios a la hora de resolver esta problemática dando un soporte bien fundamentado para el diseño y construcción automatizada de sitios Web. Para esta Línea de Investigación se propone profundizar en estos temas mediante el estudio, diseño y definición de distintos lenguajes de modelado basados en reglas que permitan distintos tipos de representación de los sitios Web así como transformaciones correctas y completas de las distintas representaciones que permitan pasar de una a otra según convenga abarcando aspectos dinámicos y estáticos. En primer lugar, esto implica lograr la identificación de fragmentos decidibles de lenguajes lógicos que permitan especificar propiedades de los sitios Web. Es necesaria la implementación de algoritmos de decisión

para verificar si se cumplen las propiedades deseadas. El trabajo sobre verificación en base a Lenguajes funcionales es el proyecto [22]. Por otro lado obtendrá un desarrollo de análisis automáticos y transformaciones que permitan optimizar, corregir e informar sobre el estado de un sitio Web en función de determinadas especificaciones de calidad para los mismos. Estas contribuciones inducen a la creación de mecanismos de certificación de sitios web. Formalizaremos un lenguaje de alto nivel capaz de analizar y diseñar sitios Web (que pueden verse como conjuntos de datos semiestructurados, por ejemplo, documentos XML/XHTML). El explosivo desarrollo de Internet y la información relacionada con las tecnologías de la información han puesto de manifiesto los problemas de la sobrecarga de información, vivimos en un entorno saturado de información, en el cual la gestión de los datos se está convirtiendo en una tarea pesada. En este escenario, la verificación y corrección de la información asume un papel crucial, que no puede ser ignorado. En base a lo expuesto, se espera como contribuciones importantes el estudio, modelado y el análisis de sitios Web.

Respecto a la Formación de Recursos Humanos, como resultado de las tareas realizadas se está desarrollando una tesis doctoral en Ciencias de la Computación y se espera delinear tesis de Maestría y tesis de Licenciatura, a medida que avance el trabajo de investigación.

Referencias

- [1] S. El-Ansary, D. Grolaux, P. Van Roy and M. Rafea: Overcoming the Multiplicity of Languages and Technologies for Web-Based Development Using a Multiparadigm Approach. Springer LNCS - Vol. 3389. pp. 113-124. ISBN 3-540-25079-4. <http://dblp.uni-trier.de>.
- [2] S. Lucas: Rewriting-Based Navigation of Web Sites: Looking for Models and Logics © Elsevier Electronic Notes in Theoretical Computer Science, 157(2):79-85, May 2006.
- [3] M. Alpuente, S. Escobar, and M. Falaschi (Eds.) First International Workshop on Automated Specification and Verification of Web Sites (WWV 2005), March 14-15, 2005 - vol. DSIC-II/03/05 - Valencia, Spain
- [4] S. Murugesan, Y. Desphande: Web Engineering. Software Engineering and Web Application Development. Springer LNCS. Vol. 2016. ISBN: 3-540-42130-0. (2001). <http://dblp.uni-trier.de>.
- [5] O. Pastor, J. Gómez, E. Insfrán and V. Pelechano: The OO-Method Approach for Information Systems Modelling: From Object-Oriented Conceptual Modeling to Automated Programming. Information Systems. Elsevier Science. Vol. 26, pp 507-534, Number 7, 2001.
- [6] O. Pastor, S. Abrahao, and J. J. Fons: An Object-Oriented approach to automate web applications development. In 2nd International Conference on Electronic Commerce and Web Technologies (EC-Web'01), Springer-Verlag, LNCS Vol: 2115. ISBN: 978-3-540-42517-5, pp. 16-28. Munich, Germany, Septiembre, 2001
- [7] O. Pastor, S. Abrahao, and J. J. Fons: Building E-Commerce Applications from Object-Oriented Conceptual Models. Newsletter of the ACM SIGecom Exchanges, volume 2.2, ACM Press, pp. 24-32. June, 2001
- [8] E. Insfrán, O. Pastor, R. Wieringa: Requirements Engineering-Based Conceptual Modelling. Requirements Engineering (RE'02). Springer-Verlag. ISSN: 0947-3602 (printed version) ISSN: 1433-010X (electronic version), pp. 61- 72. March 2002.
- [9] J. Fons, P. Valderas, O. Pastor: Specialization in Navigational Models. Argentine Conference on Computer Science and Operacional Research. Subserie ICWE, Iberoamerican Conference on Web Engineering. Volume 31 of Anales JAIIO., SantaFe, Argentina (2002). pp. 16-31 ISSN: 1666-6526.

- [10] S. Abrahao, J. Fons, M. Gonzalez and O. Pastor: Conceptual Modeling of Personalized Web Applications. In 2nd International Conference on Adaptive Hipermedia and Adaptive Web Based Systems, Malaga, Spain, 2002.
- [11] Object Management Group. Unified Modeling Language Specification Version 1.4 draft. Technical report, www.omg.org, February 2001.
- [12] N. Koch, M. Wirsing: Software Engineering for Adpatative Hypermedia Applications. In: 3rd Workshop on Adaptative Hypertext and Hypermedia (2001).
- [13] J. M. Kleinberg: Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM* 46(5):604-632, 1999.
- [14] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, A. S. Tomkins: The Web as a graph: measurements, models, and methods. In Proc. of COCOON'99, LNCS 1627: 1-17, 1999.
- [15] I. D. Baxter, F. Ricca, and P. Tonella. Web Application Transformations based on Rewrite Rules. *Information and Software Technology*, 44(13), 2002.
- [16] L. Capra, W. Emmerich, A. Finkelstein, and C. Nentwich: XLINKIT: a Consistency Checking and Smart Link Generation Service. *ACM Transactions on Internet Technology*, 2(2):151185, 2002.
- [17] T. Despeyroux and B. Trousse: Semantic Verification of Web Sites Using Natural Semantics. In Proc. of 6th Conference on Content Based Multimedia Information Access (RIAO'00), 2000.
- [18] E. Di Sciascio, F. M. Donini, M. Mongiello, and G. Piscitelli: Web Applications Design and Maintenance Using Symbolic Model Checking. In Proc. 7th European Conf. on Software Maintenance and Reengineering, page 63. IEEE Computer Society, 2003.
- [19] S. M. Easterbrook, B. Nuseibeh, and A. Russo: Leveraging Inconsistency in Software Development. *IEEE Computer*, 33(4):2429, 2000.
- [20] M. Fernandez, D. Florescu, A. Levy, and D. Suciuc: Verifying Integrity Constraints on Web Sites. In Proc. of Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99), volume 2, pages 614:619. Morgan Kaufmann, 1999.
- [21] W. Fan and L. Libkin. On XML Integrity Constraints in the Presence of DTDs. *J. ACM*, 49(3):368406, 2002.
- [22] D. Ballis and J. García Vivó: A Rule-based System for Web Site Verification. In *Proc. of 1st Int'l Workshop on Automated Specification and Verification of Web Sites (WWV'05)*, volume 157(2). ENTCS, Elsevier, 2005.
- [23] C. Kirchner, Z. Qian, P. K. Singh, and J. Stuber: Xemantics: a Rewriting Calculus-Based Semantics of XSLT. Rapport de recherche A01-R-386, LORIA, 2001