

Agentes Inteligentes y Web Semántica: Verbalización en una herramienta Web de modelado ontológico

Matías Garrido¹Germán Braun^{1,2,3}Sandra Roger¹

email: {roger,german.braun}@fi.uncoma.edu.ar

¹Grupo de Investigación en Lenguajes e Inteligencia Artificial
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

²Laboratorio de I&D en Ingeniería de Software y Sistemas de Información
Departamento de Ciencias e Ingeniería de la Computación
UNIVERSIDAD NACIONAL DEL SUR

³Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

RESUMEN

El proyecto de investigación Agentes Inteligentes y Web Semántica, financiado por la Universidad Nacional del Comahue, tiene como objetivo general la generación de conocimiento especializado en el área de agentes inteligentes y en lo referente a la representación y el uso del conocimiento en sistemas computacionales basados en la Web, es decir, lo que se ha llamado la Web Semántica.

El objetivo general del trabajo de investigación es la extensión de una herramienta de modelado ontológico, denominada crowd, mediante la verbalización de un subconjunto del lenguaje de modelado conceptual UML. Esta integración permitirá generar especificaciones en Lenguaje Natural a partir de un diagrama de clases.

Esta línea de investigación se desarrolla en forma colaborativa entre docentes-investigadores de la Universidad Nacional del Comahue y de la Universidad Nacional del Sur, en el marco de proyectos de investigación financiados por las universidades antes mencionadas.

Palabras Clave: Verbalización, Generalización de Lenguaje Natural, UML, Ontologías.

CONTEXTO

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue, en el contexto del proyecto de investigación Agentes Inteligentes y Web Semántica (04/F014), por la Universidad Nacional del Sur a través del proyecto de investigación Integración de Información y Servicios en la Web (24/N027) y por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), en el contexto de una beca interna doctoral. Los proyectos de investigación tienen una duración de cuatro años y la beca doctoral con duración de 5 años, finalizando en abril de 2019.

1. INTRODUCCIÓN

Según Christel et al. [1], la etapa inicial de un proceso de desarrollo de software es la más problemática. La complejidad radica en la correcta elicitación de los requerimientos del sistema, en su entendimiento y en la transformación de éstos en un modelo computacional.

Cualquier error cometido en estas instancias se propagará hacia las fases posteriores y su corrección incrementará los costos de desarrollo, especialmente si se identifican

durante la fase de implementación o posterior a la entrega del producto.

A medida que el proceso de desarrollo transcurre, dichos requerimientos deben ser trasladados hacia una especificación formal. Curland et al. [2] afirma que es importante que un esquema conceptual sea validado por el usuario con el objetivo de asegurar que ésta modela con precisión los aspectos relevantes del dominio. Una manera efectiva de facilitar esta validación es la traducción de tal esquema (o cualquier especificación formal) en un lenguaje claro que sea fácilmente comprensible por el experto en el dominio, quien, además, puede no poseer conocimientos en modelado de sistemas.

En esta dirección surge el concepto de verbalización, el cual consiste en traducir la semántica de una teoría lógica, es decir, las relaciones entre entidades y sus restricciones, en sentencias en lenguaje natural [3][4]. En consecuencia, esta transformación proporciona una base para la mutua comprensión entre diferentes actores acerca de la semántica que una lógica capta (ya sea un modelo conceptual, reglas de negocio o una ontología), integrando clientes/usuarios en los procesos de chequeo de consistencia cuando los cambios son realizados en el diseño o en la implementación.

La verbalización es un proceso que tiene sus bases y está comprendido dentro del campo de la Generación de Lenguaje Natural (GLN). Reiter y Dale [5] lo definen como al ámbito de la inteligencia artificial y la lingüística computacional que se ocupa de la construcción de sistemas informáticos que pueden producir textos comprensibles en español, inglés u otros lenguajes naturales de alguna representación no lingüística de información.

McDonald [6], definió a la GLN como el proceso de construir un texto en lenguaje natural para cumplir objetivos comunicativos específicos. Estos sistemas tienen el desafío no sólo de decidir qué información comunicar sino cómo transformarla, de modo que se alcance la meta de comunicación deseada.

Sin embargo, poco se ha investigado en validar los modelos de software de una manera en que los usuarios puedan comprenderlos.

Existen varias herramientas de GLN a tener en cuenta, tales como ModEx (Model Explainer [7] y GeNLangUML (*Generating Natural Language from UML*) [8]. Ambas permiten la creación y edición gráfica de modelos de software orientados a objetos (OO) [9] y diagramas de clases UML [10].

respectivamente, para generar especificaciones siendo el idioma de verbalización el inglés.

Mientras que ModEx supone que los nombres de las clases como de las relaciones de un diagrama OO son

correctos y cumplen con ciertas recomendaciones, GeNLangUML tiene la ventaja de utilizar

WordNet [11], una herramienta lingüística para validar la correctitud semántica de las sentencias generadas.

Esto implica que funcione bajo una convención de nomenclatura

más amplia para los componentes de un diagrama de clases.

Con respecto a la verbalización multilingüe, y en relación al lenguaje ORM [12], DogmaModeller [3] es una aplicación que genera automáticamente verbalizaciones en lenguaje pseudo-natural. Un aspecto a destacar de esta herramienta es su extensibilidad, es decir, la capacidad de incorporar nuevos templates de verbalización para otros idiomas.

Otra herramienta a tener en cuenta, que soporta modelos ORM de segunda generación (ORM2) [13], es NORMA (Neumont ORM Architect) [2].

Por otro lado, KeY [14] es un sistema que ayuda a comprender las restricciones OCL [15], traduciéndolas automáticamente en lenguaje natural.

La investigación desarrollada en [16] presenta un algoritmo que proporciona parafraseos en lenguaje natural para Ontologías OWL [17] en la Web Semántica (WS) [18]. Bajo la misma dirección, se encuentra NaturalOWL [19]. Generalmente los sistemas de GLN agregan el máximo de frases posibles para mejorar la legibilidad, pero NaturalOWL posee la ventaja de configurar el número máximo de frases que se desean concatenar.

2. LÍNEA DE INVESTIGACIÓN Y DESARROLLO

El proyecto de investigación Agentes Inteligentes y Web Semántica tiene como objetivo general generar conocimiento especializado en el área de agentes inteligentes y en lo referente a la representación y el uso del conocimiento en sistemas computacionales basados en la web, es decir lo que se ha llamado la Web Semántica.

Por otro lado, en el proyecto de investigación Integración de Información y Servicios en la Web se propone investigar y desarrollar metodologías y herramientas que favorezcan la interoperabilidad semántica de información y de servicios en la Web, fundamentados en los últimos avances en el área de lenguajes de representación del conocimiento, ontologías y modelado conceptual.

Ambos proyectos confluyen en la línea de investigación de este trabajo, en la que se explora entre otros, sobre temas afines a la Representación del Conocimiento, las Lógicas Descriptivas, [20], las Ontologías, la Ingeniería de Software basada en Conocimiento y la Ingeniería de Conocimiento.

En base a las investigaciones existentes hasta la fecha, se propone modificar la arquitectura de una herramienta de modelado ontológico llamada *crowd* [21], mediante la incorporación de una nueva funcionalidad que soporte verbalización multilingüe del metamodelo UML, orientado hacia un subconjunto de primitivas de los diagramas de clases. Se pretende la utilización de una representación intermedia de los diagramas de clases en sentencias concretas de lógica de primer orden (FOL) para favorecer las verbalizaciones en diversos idiomas.

Como principal novedad con respecto a las aplicaciones vigentes, se puede destacar que los usuarios podrán visualizar y editar de forma *on-line* diagramas de clases UML y posteriormente, realizar su verbalización a lenguaje natural. Esto se debe a que *crowd* es soportada por una arquitectura cliente-

servidor, lo que conlleva a que pueda ser utilizada bajo cualquier plataforma.

Otro aspecto relevante, es que la traducción a lenguaje natural está destinada a ser multilingüe, a diferencia de la mayoría de las herramientas mencionadas anteriormente, donde el lenguaje natural de salida es preferentemente el inglés.

3. RESULTADOS OBTENIDOS/ESPERADOS

Este trabajo forma parte de la propuesta presentada inicialmente en [22]. Se trabajó en el análisis y diseño de la extensión de la arquitectura de la herramienta *crowd*. Se diseñó el módulo de verbalización de dicha arquitectura y se está trabajando en la fase de implementación del mismo.

Se trabajó en la selección del subconjunto de primitivas en UML destinadas a la verbalización en una primera etapa del proceso. Además, se estableció la correspondencia directa entre estas primitivas y las sentencias de FOL.

Actualmente se está trabajando en la fase de verbalización de esas sentencias FOL. Posteriormente se prevee la extensión a otros idiomas y la correspondiente validación.

4. FORMACIÓN DE RECURSOS HUMANOS

Durante la realización de este sistema se espera lograr, como mínimo, la culminación de 2 tesis de grado dirigidas y/o codirigidas por los integrantes del proyecto. Uno de los autores de este trabajo está inscripto en el Doctorado en Ciencias de la Computación en la Universidad Nacional del Sur (beca interna doctoral CONICET).

Finalmente, es constante la búsqueda hacia la consolidación como investigadores de los miembros más recientes del grupo.

5. BIBLIOGRAFÍA

- [1] Michael Christel and Kyo Kang. Issues in requirements elicitation. Technical Report CMU/SEI-92-TR-012, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 1992.

- [2] Matthew Curland and Terry A. Halpin. The NORMA software tool for ORM 2. In Information Systems Evolution - CAiSE Forum 2010, Hammamet, Tunisia, June 7-9, 2010, Selected Extended Papers, pages 190–204, 2010.
- [3] Mustafa Jarrar. Towards Methodological Principles for Ontology Engineering. PhD thesis, Vrije Universiteit Brussel, Brussels, 5 2005.
- [4] Mustafa Jarrar, María Keet. Multilingual verbalization of ORM conceptual models and axiomatized ontologies. Vrije Universiteit Brussel. Brussels, 2006.
- [5] Reiter Euhand, Robert Dale. Building Applied Natural Language Generation Systems. Cambridge University Press. NJ, USA, 1989.
- [6] McDonald David, Gabriel Richard. Natural Language Generation Systems. 1988.
- [7] Benoit Lavoie, Owen Rambow, and Ehud Reiter. The modelexplainer, 1996.
- [8] Farid Meziane, Nikos Athanasakis, and Sophia Ananiadou. Generating natural language specifications from UML class diagrams. *Requir. Eng.*, 13(1):1–18, 2008.
- [9] Won Kim and Frederick Lochovsky. Object-oriented concepts, databases, and applications. ACM Press/Addison-Wesley Publishing Co 1989.
- [10] UML, OMG. Unified Modeling LanguageTM. 2015.
- [11] George Miller. Wordnet: a lexical database for English. ACM.1995.
- [12] Terry Halpin and Tony Morgan. Information Modeling and Relational Databases. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2 edition, 2008.
- [13] Robert Meersman and Zahir Tari. On the Move to Meaningful Internet Systems 2005: {OTM} 2005 Workshops, {OTM} Confederated International Workshops and Posters. 2005
- [14] Ahrendt, Wolfgang and Baar, Thomas and Beckert. The KeY tool. 2005.
- [15] Warmer, Jos and Kleppe, Anneke. The Object Constraint Language: Precise Modeling with UML. 1999.
- [16] Daniel Hewlett and Aditya Kalyanpur and Vladimir Kolovski and Christian Halaschek-Wiener. Effective NL Paraphrasing of Ontologies on the Semantic Web. 2005.
- [17] McGuinness, Deborah L. and van Harmelen, Frank. OWL Web Ontology Language Overview. W3C. 2004.
- [18] The semantic web. Berners-Lee, Tim and Hendler, James and Lassila, Ora and others. New York, NY, USA. 2001
- [19] Ion Androutsopoulos and Gerasimos Lampouras and Dimitrios Galanis. Generating Natural Language Descriptions from OWL Ontologies: the NaturalOWL System. 2014.
- [20] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. The Description Logic Handbook: theory, implementation, and applications. 2003.
- [21] Christian Gimenez, Germán Braun, Laura Cecchi, and Laura Fillottrani. crowd: A Tool for Conceptual Modelling assisted by Automated Reasoning - Preliminary Report. In Proc. of the 2nd Simposio Argentino de Ontologías y sus Aplicaciones (SAOA) colocated at

Jornadas Argentinas de Informática (JAIO) - to appear, 2016.

- [22]Matías Garrido, Germán Braun, Sandra Roger. Agentes Inteligentes y Web Semántica: Hacia la Verbalización de un Subconjunto de UML en una Herramienta Gráfica Web. XIX Workshop de Investigadores en Ciencias de la Computación (WICC). 2017.