

Desarrollo de API de consulta a fuentes de información en la web para Sistemas de Argumentación Rebatible

Cristian D. Pacifico, Luis I. Pelayo Novello, Leandro J. Tisocco Pertus

Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos
Av. Tavella 1425, (E3202KAC) Concordia - ER, Argentina - Tel: (0345) 423-1400
e-mail: {cripac, inovello}@fcad.uner.edu.ar, ltisocco@gmail.com

Alejandro J. Garcia, Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur
Av. Alem 1253, (B8000CPB) Bahía Blanca - BA, Argentina - Tel: (0291) 459-5135
e-mail: {ajg, grs}@cs.uns.edu.ar

Resumen

Es cada vez más necesario automatizar los procesos que integran la información que se encuentra disponible en diferentes formatos, descentralizada, con orígenes heterogéneos y administraciones diferentes. Uno de los grandes desafíos al integrar fuentes de datos heterogénea, es poder manejar la inconsistencia informacional y la posible incompletitud que puede trasladarse de las fuentes locales. Los formalismos de Argumentación son adecuados para manejar este tipo de problemas. En particular, pensar en formalismos de *argumentación rebatible* (como DeLP) para la definición y automatización de la integración de fuentes de datos heterogéneas, inconsistentes entre sí e incompletas, es una opción aceptable.

El trabajo de esta línea de investigación propone desarrollar interfaces para vincular DeLP con fuentes de datos en formatos web. En especial, fuentes de sitios web descriptas en HTML y en Servicios Web. La información obtenidas de estas fuentes, podrán ser utilizadas como *hechos* o *presunciones* en la base de conocimiento que utiliza el sistema argumentativo.

Palabras clave: Argumentación, Integración de Información, Defeasible Logic Programming, World Wide Web, Web Services

1. Contexto

Esta línea de investigación se llevará a cabo dentro del proyecto de I+D:

PID-UNER 7041 “*Desarrollo de Sistemas de Argumentación Masiva sobre Bases de Datos Federadas*”

Este proyecto está encuadrado en el convenio de colaboración entre el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) del Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur y el Área de Agentes y Sistemas Inteligentes de la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos.

2. Introducción

Es cada vez más necesario automatizar los procesos que integran la información que se encuentra disponible en diferentes formatos, descentralizada, con orígenes heterogéneos y administraciones diferentes. Desde hace décadas existen enfoques tradicionales para integrar bases de datos relacionales para conformar bases federadas. Estos enfoques proponen desde integraciones ad-hoc para constituir almacenes particulares [11, 10]; hasta la utilización de formalismos lógicos [12] y sistemas basados en marcos

[19] para describir, en forma abstracta, el marco de integración. Teniendo presente el acceso a Internet, también se crearon mecanismos para integrar información en los formatos que le son propios; por ejemplo de documentos [2], de servicios web [14], de descripciones ontologías en formalismos de web semántica [15, 13], etc.

Un gran desafío al integrar diversas fuentes de datos, es manejar eficientemente la inconsistencia informacional y la posible incompletitud que puede trasladarse de cada fuente local. Ya se sabe que los formalismos de Argumentación son adecuados para manejar estas problemáticas. En particular, pensar en formalismos de *argumentación rebatible* para la definición y automatización de la integración de fuentes de datos heterogéneas, inconsistentes entre sí e incompletas, es una opción aceptable.

La *argumentación rebatible* (*Defeasible Argumentation*) [17] es un mecanismo de razonamiento no-monótono en donde la aceptación o el rechazo de una proposición dependen de un análisis entre argumentos a favor y en contra de esa proposición [5, 16]. Usualmente es utilizada bajo diferentes formalizaciones para capturar aspectos del razonamiento del sentido común y la representación de información incompleta y potencialmente inconsistente [1, 3, 18].

La investigación propuesta aquí, se basa en la *Programación Lógica Rebatible (DeLP)* [9] siendo este es un formalismo que combina programación lógica y argumentación rebatible. Como rasgo particular, DeLP utiliza argumentación para decidir entre información contradictoria a través de un análisis dialéctico, lo que justifica la conclusiones obtenidas.

3. Línea de investigación

Esta investigación se vincula con el desarrollo de razonadores de argumentación rebatible para lograr integración de información masiva, cuyo marco teórico fue ya establecido; se ha presentado un framework (DBI-DeLP) en [6] que integra DeLP con bases de datos para lograr la ejecución de procesos de argumentación sobre fuentes de datos masivas externas y actualizables [7]. En concreto y sobre este framework se han desarrollado sistemas recomendadores [4].

El trabajo de esta línea de investigación propone *desarrollar interfaces para vincular DeLP*

con fuentes de datos en formatos web. En especial, fuentes residentes en sitios web descriptas en HTML y en Servicios Web soportados con SOAP y RESTFull. La información obtenidas de estas fuentes, podrán ser utilizadas como *hechos* o *presunciones* [8, 16] en la base de conocimiento que utiliza el sistema argumentativo.

Para llevar a cabo este objetivo, se propone desarrollar una interfaz de programación (API). Se opta por lenguajes de Programación Lógica para desarrollo de las APIs, con el ánimo de que sean integrables con DeLP. La API contendrán una estructura común que se subdivide en tres módulos:

1. Módulo **main**: contendrán los predicados/funciones que permiten recuperar, actualizar y consultar información de fuentes en la web. Estos predicados/funciones deben ser abstractos al formato de información y a los protocolos de transporte de las fuentes.
2. Módulo **sources**: contendrá los predicados/funciones que permiten definir y establecer la conexión particular con cada fuente. Para el caso de páginas web, se asocia a cada URI parametrizable, la manera de recuperación y filtrado de documentos HTML para la obtención de un dato en particular. Para el caso de servicios web, se define la estructura de la petición SOAP o RESTFull, y el filtrado de la correspondiente respuesta. Las relaciones definidas en este módulo deben configurarse para cada fuente en particular, pudiendo coexistir varias configuraciones.
3. Módulo **relations**: contendrá los predicados/funciones que vinculan relaciones semánticas con operaciones de recuperación y consulta en las fuentes. Es decir, definir la obtención de presunciones estableciendo las relaciones entre los predicados del módulo **source** para lograrlas, utilizando a éstos como funciones primitivas. Este módulo funciona como controlador de los dos anteriores.

4. Resultados y Objetivos

Actualmente, se ha implementado parte de la estructura de la API en SWI-Prolog utilizan-

do sus APIs: `http_client`, `http_sgml_plugin` y `xpath`. Como caso de testeo se han desarrollado configuraciones para tomar información de sitios de datos de películas (www.imdb.com, www.decine21.com, www.dvdempire.com y otras). Se han definido relaciones para integrar la información de directores, títulos, actores y ratings de una determinada película.

Se han definido predicados para realizar búsquedas en el sitio y obtener conjuntos resultados, por nombre de actor, director, o título. De igual forma, se desarrollo una configuración para trabajar sobre servicios web, para tratar los mismos datos.

El trabajo siguiente propone como desafío mejorar la performanse de la API, con técnicas de cacheo. Contextualizar la API en un ejemplo de argumentación real, con vinculación a DeLP. Por último, describir una metodología para describir fuentes de datos, establecer relaciones y programas DeLP en base a esta API.

5. Formación de Recursos Humanos

Esta investigación se lleva a cabo dentro del proyecto de I+D **PID-UNER 7041** “Desarrollo de Sistemas de Argumentación Masiva sobre Bases de Datos Federadas”.

Ha servido como tema de plan de tesis para 2 (dos) becas de Formación de Recursos Humanos UNER destinadas a alumnos de grado.

Está previsto que el trabajo continúe siendo la base para el plan de 1 (una) de Iniciación a la Investigación en Proyectos UNER.

La fase de desarrollo de la API, sirve como disparador para conformar temas de Tesis de grado en la Licenciatura en Sistemas de la FCAD-UNER.

Referencias

- [1] Bench-Capon, T.J.M. y P.E. Dunne: *Argumentation in artificial intelligence*. Artificial Intelligence, 171(10-15), 2007. Argumentation in Artificial Intelligence.
- [2] Bergamaschi, S., S.Castano y M.Vincini: *Semantic Integration of Semistructured and Structured Data Sources*. SIGMOD Record, 28(1), 1999.
- [3] Besnard, Philippe y Anthony Hunter: *Elements of Argumentation*. MIT Press, 2008.
- [4] Briguez, C.E., M.C. Budán, C.A.D. Deagustini, A.G. Maguitman, M. Capobianco y G.R. Simari: *Towards an Argument-based Music Recommender System*. Volumen 245 de *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2012.
- [5] Chesñevar, C.I., A.G. Maguitman y R.P. Loui: *Logical Models of Argument*. ACM Computing Surveys, 32:337–383, 2000.
- [6] Deagustini, C.A.D., S.E. Fulladoza Dalibón, S. Gottifredi y G.R. Simari: *DBI-DeLP: a framework for defeasible argumentation over databases*. XII Workshop Agentes y Sistemas Inteligentes, 2011.
- [7] Deagustini, C.A.D., S.E. Fulladoza Dalibón, S. Gottifredi, M.A. Falappa y G.R. Simari: *Consistent Query Answering Using Relational Databases through Argumentation*. En *Database and Expert Systems Applications*, volumen 7447, páginas 1–15. Springer Berlin Heidelberg, 2012.
- [8] Garcia, A., C.Chesñevar y G.Simari: *An argumentative framework for reasoning with inconsistent and incomplete information*. Workshop on practical reasoning and rationality. 13th biennial European Conference on Artificial Intelligence, 1998.
- [9] García, A.J. y G.R. Simari: *Defeasible logic programming: an argumentative approach*. Theory Pract. Log. Program., 4:95–138, 2004.
- [10] Heimbigner, D. y D. McLeod: *A federated architecture for information management*. ACM Trans. Inf. Syst., 3(3), 1985.
- [11] Inmon, W. H.: *Building the data warehouse (2nd ed.)*. John Wiley & Sons, Inc., 1996.
- [12] Lakshmanan, L. y N. Shiri: *A Parametric Approach to Deductive Databases with Uncertainty*. IEEE Trans. on Knowl. and Data Eng., 13(4), 2001.
- [13] McGuinness, D. L.: *Ontologies Come of Age*. En *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press, 2002.

- [14] Mcilraith, S.A., T.C. Son y Honglei Zeng: *Semantic Web services*. Intelligent Systems, IEEE, 16, 2001.
- [15] Meersman, R.: *Semantic Ontology Tools in IS Design*. En *11th International Symposium on Foundations of Intelligent Systems*. Springer-Verlag, 1999.
- [16] Nute, Donald: *Defeasible Logic*. En *Handbook of Logic in AI and Logic Programming*. Oxford University Press, 2001.
- [17] Pollock, John L.: *Defeasible reasoning*. Cognitive Science, 11:481-518, 1987.
- [18] Rahwan, Iyad y Guillermo R. Simari: *Argumentation in Artificial Intelligence*. Springer, 2009.
- [19] S.Heymans et al.: *Ontology Reasoning with Large Data Repositories*. En *Ontology Management*, volumen 7, páginas 89-128. Springer US, 2008.