

# Herramientas para representación y razonamiento con preferencias bajo incertidumbre

Fabio R. Gallo    Marcelo A. Falappa    Gerardo I. Simari

Inst. de Cs. e Ing. de la Computación (Universidad Nacional del Sur-CONICET)  
Av. Alem 1253, (8000) Bahía Blanca, Argentina  
{fabio.gallo,mfalappa,gis}@cs.uns.edu.ar

## Resumen

El continuo crecimiento en variedad y cantidad de actividades *online* que las personas realizan a través de Internet ha puesto en evidencia que existe una gran cantidad de potencial conocimiento disponible para quien esté dispuesto a analizarlo y tenga las herramientas apropiadas. Consideramos que este potencial no está siendo totalmente aprovechado debido a que las herramientas actuales de representación de conocimiento y razonamiento no están acompañando adecuadamente esta evolución. En este trabajo presentamos un análisis de los desarrollos previos que deben ser adaptados y combinados para acompañar a la evolución de Web. La presente línea de investigación involucra la combinación de un conjunto de áreas que hasta ahora han evolucionado prácticamente de manera independiente: Big Data, Web Semántica, Lenguajes ontológicos extendidos y Razonamiento acerca de preferencias.

**Palabras clave:** Representación de conocimiento y razonamiento, Razonamiento bajo incertidumbre probabilística, Bases de datos, Lenguajes de ontología.

## 1. Contexto

La presente línea de investigación se encuentra inserta en el marco del proyecto PGI 24/N035 “*Argumentación y Dinámica de Creencias para mejorar las capacidades de razonamiento y representación de conocimiento de Sistemas Multi-Agente*”, financiado por la Universidad Nacional del Sur, y el proyecto PIP-CONICET 112-201101-01000 “*Combinación de Revisión de Creencias y Argumentación para mejorar las capacidades de Razonamiento y modelado de la Dinámica de Conocimiento en Sistemas Multi-agente*”, financiado por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), ambos llevado a cabo dentro del Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur.

## 2. Introducción

Recientemente se ha producido una verdadera explosión en la cantidad y tipo de información que está disponible a cualquier persona con una conexión a la red Internet. Esta explosión tiene sus raíces en la World Wide Web, la cual ha revolucionado las aplicaciones de Internet, permitiéndoles a los usuarios vincular los recursos unos con otros, y organizar fácilmente el material que ellos desean hacer público. La segunda revolución llegó con el advenimiento de las aplicaciones Web, en la cual los usuarios producen su propio material, tales como en blogs o foros donde los usuarios comparten información que va desde texto plano a fotografías, audio y video. Por ello los usuarios son participantes centrales en la creación de contenido (tales como redes sociales, blogs y foros), lo cual ha llevado a la adopción del nombre Web 2.0 para señalar que ha habido una evolución en el concepto originalmente propuesto.

En los últimos años la Web tuvo una vez más un paso evolutivo, a lo que muchos se refieren como la Web Social (o Web 3.0), donde los usuarios y las relaciones entre ellos son los principales actores. Otro aspecto revolucionario que aparece en la Web 3.0 es que los datos e información básica son ahora también producidas automáticamente por máquinas; ejemplos de esto son la salida de datos por los sensores alojados en la mayoría de los teléfonos inteligentes, o por los hogares inteligentes que lentamente se están volviendo más presentes. Es en estas dos últimas “evoluciones” que el tema de los grandes datos (Big Data, en inglés) toma un rol central, dado que cualquier propuesta para el procesamiento de este tipo de información deberá ser capaz de manejar el torrente de datos que generan constantemente los usuarios y las máquinas a la par.

Más allá de los nombres adoptados y la discrepancia que puede existir en su uso, es claro que los desarrollos tecnológicos en el manejo de información ha llegado a un punto tal que es necesario investigar cómo aprovechar las capacidades ofrecidas por estos desarrollos de la mejor manera posible. En particular, nuestra investigación se enfoca en el desarrollo de mecanismos de representación de conocimiento. En el entorno de la Web Semántica y sus derivados, los mecanismos básicos utilizados para este propósito en general se denominan *len-*

*guajes ontológicos*, los cuales suelen ser adaptaciones de la lógica de primer orden para cumplir con objetivos de expresividad y tratabilidad computacional [7, 1].

Toda la información de usuarios disponible es una buena fuente de preferencias que los mismos expresan a su manera sobre, virtualmente, cualquier tópico; por ejemplo, preferencias sobre alimentos, indumentaria, ocio, política, etc. Dado que las preferencias pueden ser vistas como un tipo de conocimiento ontológico, y que aunque se elija representar las preferencias en forma diferenciada para poder razonar con ellas, los lenguajes ontológicos extendidos deberían ser investigados y adaptados para abordar esas cuestiones. Una manera de trabajar con preferencias es establecerlas como tuplas las cuales se corresponden con átomos básicos desde el punto de vista de la lógica clásica y lenguajes ontológicos. Esta es la perspectiva mayormente adoptada en la comunidad de Bases de Datos y, en general, especifican cómo deben ordenarse los resultados de una consulta. Esta perspectiva ha sido estudiada por casi tres décadas: a partir del trabajo fundacional de [16], en [26] se presenta un resumen de los trabajos más importantes en esta línea. En forma más o menos paralela, también se ha estado investigando la intersección entre las áreas de Bases de Datos y de Representación de Conocimiento y Razonamiento, tales como en los programas lógicos de preferencia de [15], la incorporación de preferencias en formalismos como programación con la semántica de Answer Sets [6], y mecanismos de respuesta de consultas top-k en *lenguajes ontológicos* [14]. Además de los trabajos de [?, 19] que representan una aproximación inicial a problemas específicos, no se tiene conocimiento de trabajos que combinen razonamiento probabilístico con representaciones lógicas como las propuestas aquí para razonar acerca de preferencias. A su vez, es esencial llevar a cabo una evaluación empírica extensa de los resultados obtenidos.

Por último, un proceso de razonamiento que hasta ahora no ha sido aplicado en el razonamiento acerca de preferencias es el de *argumentación*. El proceso de argumentación refleja una forma de razonamiento donde tanto la conclusión como la forma de llegar a ella pueden ser cuestionadas. Este área de estudio es de especial interés en el ámbito de la Inteligencia Artificial, principalmente porque permite razonar con información incompleta e incierta, y permite manejar inconsistencias en los sistemas basados en conocimiento. Por lo tanto, este tipo de razonamiento puede ser útil para nuestros objetivos. Como punto de partida en esta investigación se considerarán tanto los sistemas de argumentación abstractos [9] como los sistemas concretos basados en reglas. Algunos ejemplos de sistemas que han sido desarrollados en las últimas décadas son: Defeasible Logic Programming (DeLP) [13], ABA [10] y ASPIC+ [21].

### 3. Líneas de Investigación, Desarrollo e Innovación

A continuación haremos un breve relevamiento de los ejes de investigación, enfocándonos en los trabajos que comparten en mayor medida nuestras motivaciones.

#### Big Data, Web Semántica y sus evoluciones

La Web Semántica [2] es un proyecto ambicioso que fue propuesto hace poco más de una década para la integración de conocimiento en la red de documentos disponibles en Internet, comúnmente llamada la *Web*. La idea principal es lograr que el contenido que existe en dichos documentos sea accesible y procesable computacionalmente para que pueda ser compartida y reusada en forma eficiente en tareas que, en general, pueden ser resueltas con mayor o menor facilidad por humanos pero no por computadoras. Por ejemplo, con un poco de trabajo un ser humano es capaz de encontrar y reservar un hotel que cuente con las prestaciones que necesita y prefiere; sin embargo, un sistema computacional se encuentra con muchos obstáculos, tales como: (1) la existencia de varios sitios Web que ofrecen el servicio de reserva de hoteles pero usan formatos de datos muy diferentes; (2) incertidumbre en los datos, tales como inconsistencia (por ejemplo, en una parte del sitio afirma que el desayuno es gratis mientras que en otra tiene un precio asociado) o incompletitud (por ejemplo, el hotel afirma disponer de conexión a Internet, pero en ninguna parte dice si es inalámbrica o por cable); y (3) el sistema automatizado no conoce las preferencias del usuario más allá de las básicas que fueron tomadas como parte de la consulta. Por lo tanto, un sistema computacional sería incapaz de decidir entre opciones que parecen ser equivalentes pero que el usuario ordenaría de una manera específica sin dudar. Por ejemplo, si el usuario quiere un hotel cerca del centro de Mar del Plata con Internet gratis, tal vez existan muchas opciones; sin embargo, dada la información de ubicación de cada uno, el usuario elegiría aquél más cercano a la playa porque esa es una de sus actividades preferidas.

#### Razonamiento acerca de preferencias

El estudio de la representación y el razonamiento acerca de preferencias se ha llevado a cabo en áreas tan dispares como filosofía, lógica matemática y economía, entre otras. En la tradición adoptada en filosofía, en general, las preferencias se expresan sobre “desenlaces” (comúnmente denominados *outcomes*, en Inglés), los cuales muchas veces se corresponden con asignaciones de valores de verdad a las variables de un conjunto de fórmulas en lógica de primer orden. En este marco, el trabajo de [3] apunta a cerrar la brecha entre varios formalismos propuestos en la comunidad de Inteligencia Artificial, tales como CP-Nets [5], y aquellos estudiados en Filosofía. Desde el punto de vista del razonamiento bajo incertidumbre, no existe literatura extensa que se enfoque en la

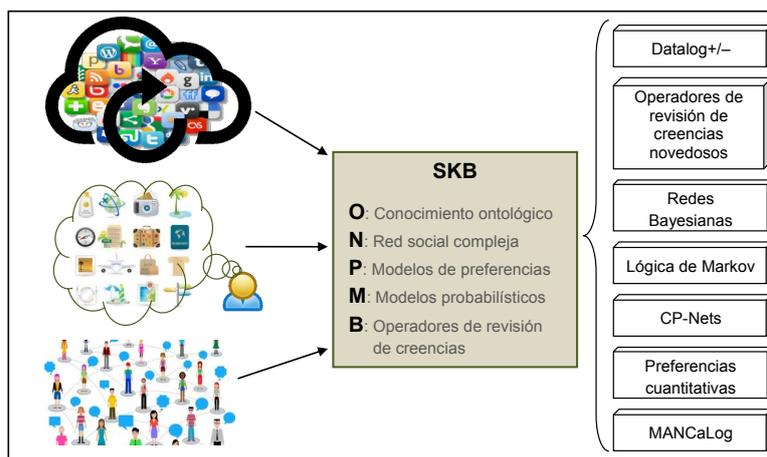


Figura 1: Las *bases de conocimiento social* (SKBs) se construyen a partir de información de la Web, como la que surge de la interacción de los usuarios con servicios *online* y redes sociales. Las componentes individuales se modelan con diferentes tipos de formalismos propuestos en la literatura para resolver problemas más específicos.

combinación de ambos; las extensiones probabilísticas de CP-Nets estudiadas en [4, 8] son dos ejemplos recientes.

### Lenguajes ontológicos extendidos

Los lenguajes ontológicos, sistemas basados en reglas, y la integración de ambos es un tema central para desarrollos en la Web Semántica. Si bien existen muchas propuestas para esta integración, tales como acoplamientos débiles, fuertes e híbridos, como así también en la generalización de lenguajes ontológicos para expresar reglas, la familia de lenguajes Datalog+/- [7] es la primera propuesta que generaliza las reglas y dependencias de bases de datos para la expresión de axiomas ontológicos. El desarrollo de dicha familia, que comenzó hace pocos años, se está produciendo en un momento en el que existe gran interés en la comunidad de la Web Semántica en disponer de formalismos altamente escalables. Esta falta de desarrollo hace que exista un vacío en la disponibilidad de formalismos con la capacidad de expresar conocimiento ontológico, razonamiento bajo incertidumbre y escalabilidad similar a la lograda en los sistemas de bases de datos modernos. Las propuestas de [18, 17] dan algunos pasos en esta dirección, pero existe mucho camino que recorrer en cuanto a la comprensión de cómo lograr una integración balanceada de dichas funcionalidades; a su vez, hace falta llevar a cabo un programa de implementación y evaluación empírica riguroso, con datos tanto reales como sintéticos, para complementar las propuestas teóricas.

## 4. Resultados y objetivos

El objetivo es investigar los aspectos computacionales del razonamiento acerca de preferencias (de entidades que pueden ser personas o grupos enteros) bajo la existencia de incertidumbre, con un enfoque particular en la tratabilidad computacional para poder aplicar los resultados a grandes datos (Big Data, en inglés).

El objetivo particular posee dos partes: (i) El desarrollo teórico de formalismos con la capacidad de representar preferencias bajo incertidumbre, incluyendo el análisis de la complejidad computacional y los factores que influyen en ella; y (ii) El análisis empírico de dichos formalismos utilizando tanto datos reales como sintéticos. El primer objetivo permitirá la identificación de aquellos formalismos que, al menos en teoría, exhiban un buen balance entre expresividad y tratabilidad computacional. Como contrapartida, el segundo objetivo permitirá confirmar que la teoría condice con la práctica y/o identificar cuáles aspectos necesitan ser ajustados. Por lo tanto, ambos objetivos deberán ser desarrollados en forma concurrente para que los resultados de uno puedan informar a los del otro.

Como punto de partida para esta investigación, actualmente se estudian los elementos básicos que serán necesarios combinar, a saber: lenguajes ontológicos, tales como los que componen la familia de Datalog+/- [7], modelos probabilísticos como las Redes Bayesianas [20] y la Lógica de Markov [22], y métodos para el manejo de preferencias, tales como los que se analizan en [26].

Como primer abordaje a las problemáticas aquí mencionadas, recientemente elaboramos una propuesta en la cual se propone una desiderata (una lista de características deseables) para el desarrollo de lo que denominamos *Bases de Conocimiento Social* (SKB, por sus siglas en inglés) [11]. El objetivo a largo plazo de ese trabajo es derivar un framework y una metodología similar a Ontology Based Data Access (OBDA) [25] que esté especializado en los aspectos sociales únicos mencionados anteriormente. Esta desiderata está influenciada por propuestas anteriores en la literatura de redes complejas [23, 24], la cual establece un conjunto de criterios deseables para modelos en cascada. La figura 1 muestra un esquema de alto nivel de cómo se organiza y compone una SKB.

Esta línea de investigación está también estrechamente relacionada con otros proyectos que se están llevando a cabo en el LIDIA; sus investigadores han sido partici-

pantes centrales en diferentes proyectos de investigación sobre revisión de creencias, sistemas multi-agente, lenguajes ontológicos para razonar bajo incertidumbre, razonamiento en redes complejas, robótica cognitiva, y razonamiento acerca del comportamiento de adversarios en el marco de proyectos de modelamiento cultural computacional. Dentro de estos proyectos se desarrolló una implementación del lenguaje de argumentación estructurada *DeLP* (por sus siglas en Inglés de *Defeasible Logic Programming*) [13] y hoy cuenta con una versión disponible que permite el uso del sistema a través de la Web <sup>1</sup>. Se ha desarrollado también una versión llamada *DeLP-Server* que provee un servicio de razonamiento para sistemas multi-agentes donde diferentes agentes en host remotos pueden hacer uso de este servicio de razonamiento argumentativo [12]. Estos sistemas serán utilizados a su vez en la presente línea de investigación.

## 5. Formación de Recursos Humanos

Dentro de esta línea de investigación se lleva a cabo la tesis de Doctor en Ciencias de la Computación de Fabio R. Gallo, bajo la dirección de Marcelo A. Falappa y Gerardo I. Simari, en desarrollo dentro del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) perteneciente al Instituto de Ciencias e Ingeniería de la Computación (ICIC), instituto de doble dependencia de la Universidad Nacional del Sur y CONICET. Actualmente, el LIDIA cuenta con investigadores, becarios y estudiantes de posgrado trabajando intensamente en las áreas de Razonamiento bajo Incertidumbre e Inconsistencia, Web Semántica, Razonamiento sobre Preferencias, Robótica Cognitiva, Argumentación Rebatible, Revisión de Creencias y Sistemas Multi-agente.

## Referencias

- [1] BAADER, F. *The description logic handbook: Theory, implementation and applications*. Cambridge university press, 2003.
- [2] BERNERS-LEE, T., HENDLER, J., LASSILA, O., ET AL. The semantic web. *Scientific american* 284, 5 (2001), 28–37.
- [3] BIENVENU, M., LANG, J., AND WILSON, N. From preference logics to preference languages, and back. In *KR*, pp. 214–224.
- [4] BIGOT, D., FARGIER, H., MENGIN, J., AND ZANUTTINI, B. Probabilistic conditional preference networks. *arXiv preprint arXiv:1309.6817*, 72–81.
- [5] BOUTILIER, C., BRAFMAN, R. I., DOMSHLAK, C., HOOS, H. H., AND POOLE, D. Cp-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *J. Artif. Intell. Res. (JAIR)* 21 (2004), 135–191.
- [6] BREWKA, G. Preferences, contexts and answer sets. In *ICLP* (2007), p. 22.
- [7] CALÌ, A., GOTTLOB, G., AND LUKASIEWICZ, T. A general datalog-based framework for tractable query answering over ontologies. *Web Semantics: Science, Services and Agents on the World Wide Web* 14 (2012), 57–83.
- [8] CORNELIO, C., GOLDSMITH, J., MATTEI, N., ROSSI, F., AND VENABLE, K. B. Updates and uncertainty in cp-nets. In *AI 2013: Advances in Artificial Intelligence*. Springer, 2013, pp. 301–312.
- [9] DUNG, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77, 2 (1995), 321–358.
- [10] DUNG, P. M., KOWALSKI, R. A., AND TONI, F. Assumption-based argumentation. In *Argumentation in artificial intelligence*. Springer, 2009, pp. 199–218.
- [11] GALLO, F. R., ABAD SANTOS, N., SIMARI, G. I., AND FALAPPA, M. A. A desiderata for modeling and reasoning with social knowledge. In *XXI Congreso Argentino de Ciencias de la Computación (Junín, 2015)* (2015).
- [12] GARCÍA, A. J., ROTSTEIN, N. D., TUCAT, M., AND SIMARI, G. R. An argumentative reasoning service for deliberative agents. In *Knowledge Science, Engineering and Management, Second International Conference, KSEM 2007, Melbourne, Australia, November 28-30, 2007, Proceedings* (2007), pp. 128–139.
- [13] GARCÍA, A. J., AND SIMARI, G. R. Defeasible logic programming: An argumentative approach. *Theory and practice of logic programming* 4, 1+ 2 (2004), 95–138.
- [14] GOTTLOB, G., LUKASIEWICZ, T., MARTINEZ, M. V., AND SIMARI, G. I. Query answering under probabilistic uncertainty in datalog+/- ontologies. *Annals of Mathematics and Artificial Intelligence* 69, 1 (2013), 37–72.
- [15] GOVINDARAJAN, K., JAYARAMAN, B., AND MANTHA, S. Preference logic programming. In *Logic Programming, Proceedings of the Twelfth International Conference on Logic Programming, Tokyo, Japan, June 13-16, 1995* (1995), pp. 731–745.
- [16] LACROIX, M., AND LAVENCY, P. Preferences; putting more knowledge into queries. In *VLDB* (1987), vol. 87, pp. 1–4.

<sup>1</sup>[http://lidia.cs.uns.edu.ar/delp\\_client/](http://lidia.cs.uns.edu.ar/delp_client/)

- [17] LUKASIEWICZ, T., MARTINEZ, M. V., ORSI, G., AND SIMARI, G. I. Heuristic ranking in tightly coupled probabilistic description logics. *arXiv preprint arXiv:1210.4894* (2012).
- [18] LUKASIEWICZ, T., MARTINEZ, M. V., AND SIMARI, G. I. Consistent answers in probabilistic datalog+/- ontologies. In *Web Reasoning and Rule Systems - 6th International Conference, RR 2012, Vienna, Austria, September 10-12, 2012. Proceedings* (2012), pp. 156–171.
- [19] LUKASIEWICZ, T., MARTINEZ, M. V., AND SIMARI, G. I. Probabilistic preference logic networks. In *ECAI* (2014), pp. 561–566.
- [20] PEARL, J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1998.
- [21] PRAKKEN, H. An abstract framework for argumentation with structured arguments. *Argument and Computation* 1, 2 (2010), 93–124.
- [22] RICHARDSON, M., AND DOMINGOS, P. Markov logic networks. *Machine learning* 62, 1-2 (2006), 107–136.
- [23] SHAKARIAN, P., SIMARI, G. I., AND CALLAHAN, D. Reasoning about complex networks: A logic programming approach. *TPLP* 13, 4-5-Online-Supplement (2013).
- [24] SHAKARIAN, P., SIMARI, G. I., AND SCHROEDER, R. MANCaLog: A logic for multi-attribute network cascades. In *Proc. of AAMAS-2013* (2013).
- [25] SPANOS, D.-E., STAVROU, P., AND MITROU, N. Bringing relational databases into the semantic web: A survey. *Semantic Web* 3, 2 (2012), 169–209.
- [26] STEFANIDIS, K., KOUTRIKA, G., AND PITOURA, E. A survey on representation, composition and application of preferences in database systems. *ACM Transactions on Database Systems (TODS)* 36, 3 (2011), 1–45.